Novel Algorithms for Lp-Quasi-Norm Principal-Component Analysis

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 consider outlier-resistant Lp-quasi-norm ($p \leq 1$) Principal-Component Analysis (Lp-PCA) of a D-by-N matrix. It was recently shown that Lp-PCA ($p \leq 1$) admits an exact solution by means of combinatorial optimization with computational cost exponential in N. To date, apart from the exact solution to Lp-PCA ($p \leq 1$), there exists no converging algorithm of lower cost that approximates its exact solution. In this work, we (i) propose a novel and converging algorithm that approximates the exact solution to Lp-PCA with significantly lower computational cost than that of the exact solver, (ii) conduct formal complexity and convergence analyses, and (iii) propose a multi-component solver based on subspace-deflation. Numerical studies on matrix reconstruction and medical-data classification illustrate the outlier resistance of Lp-PCA.**

***Index Terms*—Principal-Component Analysis, PCA, L1-PCA, Lp-norm, quasi-norm, Lp-quasi-norm, outliers, robustness.**

I. INTRODUCTION

Principal-Component Analysis (PCA) is a standard method for the analysis of matrix data with applications in the fields of machine learning, wireless communications, computer vision, and neuroscience, to name a few [1], and is commonly used for compression, denoising, classification, and pattern recognition. From an optimization standpoint, standard PCA is formulated as a L2-norm error (Euclidean distance) minimization, or, equivalently, a L2-norm projection maximization problem. Its solution is computed by means of Singular-Value Decomposition (SVD) of the processed matrix [2]. Regretfully, researchers have long noticed that standard PCA is highly sensitive against outliers in the processed matrix [3, 4]. Outliers are erroneous points that lie far away from the nominal subspace on which the points of the processed matrix are expected to lie on. Outliers often appear in modern datasets due to data storage/transfer errors, faulty sensors, or deliberate data contamination in adversarial environments [5]. Accordingly, the performance of applications which rely on PCA is compromised, even when a small fraction of data points are outlier corrupted [6]. This sensitivity of standard PCA against outliers can be traced back to its L2-norm based formulation, which places squared emphasis on each data point of the processed matrix, benefiting high-magnitude/peripheral points. To remedy the impact of outliers, researchers have proposed multiple *robust* PCA formulations [6, 7]. For instance, [8] and [9] consider L1-norm and Lp-norm, respectively, based residual-error minimization formulations and approximate a solution by means of alternating-optimization. In contrast, authors in [10] consider a L1-norm based projection maximization approach which derives by simple substitution of the outlier-responsive L2-norm in the standard projection-maximization PCA formulation by the more robust L1-norm which places linear emphasis on each data point of the processed matrix. This change in norm results in the popular L1-norm PCA (L1-PCA) formulation. L1-PCA has been solved both exactly [10] and approximately [11, 12]. Moreover, stochastic, adaptive, incremental, and complex-valued L1-PCA solvers have been presented in [13–18]. Image recovery, video surveillance, and visual tracking [19–22] are just a few applications in which L1-PCA has been successfully employed. It has been well documented that L1-PCA exhibits similar performance to PCA when the processed data are nominal while it exhibits strong resistance against outliers. Following the paradigm of L1-PCA, researchers proposed robust reformulations of popular multi-way (tensor) decompositions–e.g., L1-Tucker [23–30] and L1-Rescal [31]. Similar to L1-PCA, L1-norm formulations of tensor decompositions exhibit similar performance to their L2-norm based counterparts when the processed data are nominal and sturdy resistance against outliers. Apart from $p = 2$ (standard PCA) and $p = 1$ (L1-PCA), Lp-quasi-norm PCA (Lp-PCA) was recently solved exactly for any $p \leq 1$. It was shown that, for any matrix $\mathbf{X}$ of size D-by-N, Lp-PCA can be cast to combinatorial optimization and solved exactly with cost exponential in N [32]. To date, apart from the (computationally expensive) exact solution to Lp-PCA, there exists no converging approximate algorithm that approximates the exact solutuion to Lp-PCA ($p \leq 1$).

In this work, we propose a novel iterative procedure which approximates the exact solution to Lp-PCA ($p \leq 1$) with significantly lower than exponential computational cost and convergence guarantees. The proposed algorithm is accompanied by formal complexity and convergence analyses. Finally, we propose a multi-component solver which employs a subspace-deflation approach. Our experimental studies on matrix reconstruction and medical-data classification illustrate the outlier resistance of Lp-PCA.

*Corresponding author.

This research is supported in part by the U.S. National Science Foundation, under grant OAC-1808582, and the U.S. Air Force Office of Scientific Research under the Young Investigator Research Program (YIP).

II. PROBLEM STATEMENT

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

$$\max_{\mathbf{q} \in \mathbb{S}_D} \left\| \mathbf{X}^\top \mathbf{q} \right\|_p^p, \tag{1}$$

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 $\mathbb{S}_D := \{\mathbf{q} \in \mathbb{R}^D : \|\mathbf{q}\|_2 = 1\}$. First, we observe that, for $p \geq 1$, $\|\cdot\|_p$ defines a proper norm the definition of which follows [33].

Definition 1. *A norm on $\mathbb{R}^D$ is any function $f : \mathbb{R}^D \to \mathbb{R}_+$ which, for any $\mathbf{x} \in \mathbb{R}^D$, $\mathbf{y} \in \mathbb{R}^D$, and $\alpha \in \mathbb{R}$, satisfies the following properties:*

1) $f(\alpha \mathbf{x}) = |\alpha| f(\mathbf{x})$ *(absolutely homogeneous).*
2) *If $f(\mathbf{x}) = 0$, then $\mathbf{x} = \mathbf{0}_D$ (zero-vector).*
3) $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ *(triangle inequality).*

In contrast, for $p < 1$, $\|\cdot\|_p$ violates the triangle inequality property and, thus, it's not a proper norm. However, $\|\cdot\|_p$ ($p \leq 1$) is a quasi-norm–i.e., it satisfies 1), 2), and the inequality

$$f(\mathbf{x} + \mathbf{y}) \leq C \left(f(\mathbf{x}) + f(\mathbf{y}) \right), \tag{2}$$

for some constant $C \geq 1$. In fact, one can show that $\|\cdot\|_p$ satisfies (2) with $C = 2^{\frac{1-p}{p}}$ [34]. Trivially, every proper norm is also a quasi-norm that satisfies (2) with $C = 1$. Compactly, for $p > 0$, $\|\cdot\|_p$ defines a quasi-norm that satisfies (2) with $C = \max\{1, 2^{\frac{1-p}{p}}\}$. In view of the above, in the special case that $p \leq 1$, (1) reduces to a Lp-quasi-norm maximization problem. To date, Lp-PCA has been solved exactly only for $p = 2$, $p = 1$, and $p \leq 1$. For $p > 1$, approximate converging solvers have been proposed in [35] but the exact solution remains unknown. For $p < 1$, there exists no approximate algorithm with convergence guarantees. In the sequel, we briefly review the special cases of $p = 1$ (L1-PCA) and Lp-PCA ($p \leq 1$). Then, we present novel approximate algorithms for Lp-PCA ($p \leq 1$) which approximate the exact solution to Lp-PCA with low computational cost and convergence guarantees.

A. Lp-PCA ($p \leq 1$) BACKGROUND

For $p = 1$, (1) simplifies to the L1-PCA formulation $\max_{\mathbf{q} \in \mathbb{S}_D} \left\| \mathbf{X}^\top \mathbf{q} \right\|_1$. L1-PCA was solved exactly in [10], where it was proven that if $\mathbf{b}_{\text{opt}}$ is a solution to $\max_{\mathbf{b} \in \{\pm 1\}^N} \|\mathbf{X}\mathbf{b}\|_2$, then, $\mathbf{q}_{\text{L1}} = \omega(\mathbf{X}\mathbf{b}_{\text{opt}})$, solves L1-PCA exactly, where $\omega(\mathbf{x}) := \mathbf{x}\|\mathbf{x}\|_2^{-1}$ for any $\mathbf{x} \in \mathbb{R}^D$. Intelligent algorithms can solve L1-PCA with computational cost $\mathcal{O}(N^K)$ [10]–i.e., polynomial in N. [11] offered an efficient solver that approximates $\mathbf{q}_{\text{L1}}$ with cost $\mathcal{O}(DN \min\{D, N\} + N^2 K)$, similar to the cost of standard PCA (SVD). For $p \leq 1$, authors in [32] first observed that $\max_{\mathbf{q} \in \mathbb{S}_D} \left\| \mathbf{X}^\top \mathbf{q} \right\|_p^p = \max_{\mathbf{q} \in \mathbb{B}_D} \left\| \mathbf{X}^\top \mathbf{q} \right\|_p^p$, where $\mathbb{B}_D = \{\mathbf{q} \in \mathbb{R}^D : \|\mathbf{q}\|_2 \leq 1\}$. That is, maximization can be equivalently pursued over the unit-radius hyperball in $\mathbb{R}^D$. Then, they noticed that $\mathbb{B}_D$ can be partitioned into a finite number (which depends on N) of non-overlapping sets which, in turn, implies that the original (non-convex) problem of interest can be partitioned in a finite-number of convex subproblems, each of which admits a solution. Moreover, the global solution to the original problem coincides with the solution of one of the convex subproblem instances. Formally, it was shown that $\mathbb{B}_D = \cup_{\mathbf{b} \in \mathcal{B}_N} \mathcal{C}(\mathbf{b})$, where $\mathcal{B}_N = \{\pm 1\}^N$, for every $\mathbf{b} \in \mathcal{B}_N$ $\mathcal{C}(\mathbf{b}) := \{\mathbf{q} \in \mathbb{B}_D : \text{sgn}(\mathbf{X}^\top \mathbf{q}) = \mathbf{b}\}$, $\text{sgn}(\mathbf{x}) = [\text{sgn}([\mathbf{x}]_1), \ldots, \text{sgn}([\mathbf{x}]_D)]^\top$, $\text{sgn}(\alpha) = +1$ if $\alpha > 0$ and $\text{sgn}(\alpha) = -1$ if $\alpha < 0$. Without loss of generality, $\text{sgn}(\alpha)$ is set to $+1$ when $\alpha = 0$. Importantly, for every $\mathbf{b}$, $\mathcal{C}(\mathbf{b})$ is equivalently expressed as $\mathcal{C}(\mathbf{b}) = \{\mathbf{q} \in \mathbb{B}_D : [\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q} \geq 0 \forall n \in [N]\}$. In view of the above, it follows that $\max_{\mathbf{q} \in \mathbb{B}_D} \left\| \mathbf{X}^\top \mathbf{q} \right\|_p^p = \max_{\mathbf{q} \in \cup_{\mathbf{b} \in \mathcal{B}_N} \mathcal{C}(\mathbf{b})} \left\| \mathbf{X}^\top \mathbf{q} \right\|_p^p$. For any fixed $\mathbf{b} \in \mathcal{B}_N$, the following hold: $\mathcal{C}(\mathbf{b})$ is a convex set and $\|\mathbf{X}^\top \mathbf{q}\|_p^p = \sum_{n \in [N]} |\mathbf{x}_n^\top \mathbf{q}|^p = \sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q})^p$ is concave with respect to $\mathbf{q}$ when $\forall n \in [N]$ $[\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q} \geq 0$. The following Proposition 1 summarizes the exact solution to (1) for any $p \leq 1$, first presented in [32].

Proposition 1. *The exact solution to Lp-PCA in* (1) *can be computed by an exhaustive search over $\mathbf{b} \in \mathcal{B}_N$. That is, for each fixed $\mathbf{b} \in \mathcal{B}_N$, one needs to solve the convex subproblem*

$$\min_{\mathbf{q} \in \mathcal{C}(\mathbf{b})} -\sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q})^p \equiv \min_{\substack{\mathbf{q} \in \mathbb{B}_D \\ [\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q} \geq 0 \forall n}} -\sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q})^p. \tag{3}$$

The global solution to Lp-PCA coincides with the solution of the subproblem that minimizes the objective in (3).

The optimization problem in (3) can be solved efficiently with a primal-dual interior-point solver based on Newton's method [36, 37] with about cubic cost in D, N. Moreover, (3) can be solved with publicly available software like CVX [38, 39]. A CVX code snippet for solving (3) is offered in [32].

III. PROPOSED ALGORITHMS

Similar to Lp-PCA, L1-PCA and rank-1 L1-Tucker2 have been solved exactly by means of combinatorial optimization. To avoid a costly exhaustive search over $\mathcal{B}_N$, iterative approximate solvers that conduct optimal single bit-flips per iteration have been proposed for L1-PCA [11] and L1-Tucker2 [25]. Motivated by these works, in this work, we propose Lp-bit-flipping (Lp-BF) algorithm, a converging algorithm that approximates the exact solution to Lp-PCA by a search via optimal single bit flips per iteration. The proposed algorithm is described as follows. First, we initialize at an antipodal binary vector $\mathbf{b}_0 \in \mathcal{B}_N$ such that $\mathcal{C}(\mathbf{b}_0) \neq \emptyset$ (e.g., $\mathbf{b}_0 = \text{sgn}(\mathbf{X}^\top \mathbf{q})$ where $\mathbf{q}$ is arbitrary). Next, at the $(t \geq 1)$-th iteration, we search for the entry of $\mathbf{b}_{t-1}$ the negation/flipping of which offers the maximum increase in the metric of (3) (or equivalently, (1)). For instance, consider $m \in [N]$ and

$$\mathbf{b}_t'(m) = \mathbf{b}_{t-1} - 2[\mathbf{b}_{t-1}]_m \mathbf{e}_{m,N}, \tag{4}$$

where $\mathbf{e}_{m,N}$ is the m-th column of the size-N identity matrix $\mathbf{I}_N$. At the t-th iteration, the proposed algorithm computes

$$m^* = \arg\max_{m \in [N]} \max_{\mathcal{C}(\mathbf{b}_t'(m))} \sum_{n \in [N]} \left([\mathbf{b}_t'(m)]_n \mathbf{x}_n^\top \mathbf{q} \right)^p, \tag{5}$$

and sets $\mathbf{b}_t'^* = \mathbf{b}_{t-1} - 2[\mathbf{b}_{t-1}]_{m^*} \mathbf{e}_{m^*,N}$. Thereafter, we let $v(\mathbf{b}) = \max_{\mathcal{C}(\mathbf{b})} \sum_{n \in [N]} \left([\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q} \right)^p$ for any $\mathbf{b}$. If $v(\mathbf{b}_t'^*) > v(\mathbf{b}_{t-1})$, then we set $\mathbf{b}_t = \mathbf{b}_t'^*$, and proceed to the next iteration. In contrast, if m^* is such that $v(\mathbf{b}_t'^*) \leq v(\mathbf{b}_{t-1})$,

Algorithm 1. Proposed Lp-BF algorithm for Lp-PCA.

$\mathbf{q} \leftarrow \texttt{Lp-BF}(\mathbf{X}, p)$

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

Initialize $\mathbf{b} \in \{\pm 1\}^N$, $\mathcal{C}(\mathbf{b}) \neq \emptyset$

Until convergence/termination

  $\mathbf{q} \leftarrow \arg\min_{\mathbf{q} \in \mathcal{C}(\mathbf{b})} -\sum_{n \in [N]} ([\mathbf{b}]_n \mathbf{x}_n^\top \mathbf{q})^p$

  $v \leftarrow \|\mathbf{X}^\top \mathbf{q}\|_p^p$

  For every $m \in [N]$

    $\mathbf{b}'(m) \leftarrow \mathbf{b} - 2[\mathbf{b}_{t-1}]_m \mathbf{e}_{m,N}$

    $\mathbf{q}' \leftarrow \arg\min_{\mathbf{q} \in \mathcal{C}(\mathbf{b}'(m))} -\sum_{n \in [N]} ([\mathbf{b}'(m)]_n \mathbf{x}_n^\top \mathbf{q})^p$

    If $\|\mathbf{X}^\top \mathbf{q}'\|_p^p > v$, $v \leftarrow \|\mathbf{X}^\top \mathbf{q}'\|_p^p$, $\mathbf{b} \leftarrow \mathbf{b}'$

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

Fig. 1. Proposed Lp-BF solver for Lp-PCA.

then the iterations terminate and the algorithm returns $\mathbf{q} = \arg\max_{\mathbf{q} \in \mathcal{C}(\mathbf{b}_{t-1})} \sum_{n \in [N]} ([\mathbf{b}_{t-1}]_n \mathbf{x}_n^\top \mathbf{q})^p$, as the approximate solution to (1). The proposed Lp-BF algorithm is summarized in Fig. 1.

Convergence/termination: By definition, the above procedure increases the metric of (3) (and thus, (1)) at each iteration at which a single bit is flipped. Moreover, the metric is upper bounded by the exact solution of Lp-PCA obtainable by an exhaustive search. Since $\mathbf{b} \in \mathcal{B}_N$, the target metric takes a finite number of values, and thus, the presented algorithm will converge in a finite number of steps. In practice, we have observed that the algorithm terminates in $T \ll N$ iterations.

Complexity analysis: At every iteration, the algorithm solves N instances of the optimization problem in (3). Accordingly, the the per-iteration computation complexity of the proposed Lp-BF algorithm is N times the computation effort required for solving an instance of (3) which depends on the solver of choice–e.g., a primal-dual interior point solves (3) with about cubic complexity in D, N.

In the more general case where $1 \leq k \leq K$ Lp-PCs need to be extracted, Lp-PCA is formulated as

$$\max_{\mathbf{Q} \in \mathbb{S}_{D \times k}} \left\| \mathbf{X}^\top \mathbf{Q} \right\|_p^p, \tag{6}$$

where $\mathbb{S}_{D \times k} = \{\mathbf{Q} \in \mathbb{R}^{D \times k} \ : \ \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_k\}$. To date, there exists neither a converging approximate solver nor an exact solution for (6). In this work, we propose an approximate solution to (6) which employs a subspace deflation approach and extracts the columns of an orthonormal matrix $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ sequentially. Mathematically, the proposed algorithm sets $\mathbf{X}_1 = \mathbf{X}$ and computes $\mathbf{q}_1 \leftarrow \texttt{Lp-BF}(\mathbf{X}_1, p)$. Then, it sets $\mathbf{X}_2 = (\mathbf{I}_D - \mathbf{q}_1 \mathbf{q}_1^\top)\mathbf{X}$ and computes $\mathbf{q}_2 \leftarrow \texttt{Lp-BF}(\mathbf{X}_2, p)$. Generalizing, for the extraction of the k-th Lp-PC the proposed algorithm sets $\mathbf{X}_k = (\mathbf{I}_D - \sum_{i \in [k-1]} \mathbf{q}_i \mathbf{q}_i^\top)\mathbf{X}$ and computes $\mathbf{q}_k \leftarrow \texttt{Lp-BF}(\mathbf{X}_k, p)$. The proposed algorithm for the extraction of k components is summarized in Fig. 2. The computationally complexity of Algorithm 2 is k times the computational complexity of Algorithm 1.

Algorithm 2. Proposed Lp-BF algorithm for rank-($k \geq 1$) Lp-PCA.

$\mathbf{Q} \leftarrow \texttt{Lp-BF}(\mathbf{X}, p, k)$

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

For $i = 1$ to k

  $\mathbf{X}_i \leftarrow (\mathbf{I}_D - \sum_{j \in [i-1]} \mathbf{q}_j \mathbf{q}_j^\top)\mathbf{X}$, $\mathbf{q}_i \leftarrow \texttt{Lp-BF}(\mathbf{X}_i, p)$

Return: $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k] \in \mathbb{R}^{D \times k}$

Fig. 2. Proposed Lp-BF solver for rank-($k \geq 1$) Lp-PCA.

IV. Numerical studies

We commence with an empirical study on the performance degradation ratio attained by the solution of the proposed Lp-BF algorithm in Fig. 1 with respect to the exact solution of Lp-PCA. We consider $\mathbf{X}_r = \mathbf{q}\mathbf{v}^\top + \mathbf{N}_r$, where $\|\mathbf{q}\|_2 = \|\mathbf{v}\|_2 = 1$, $\mathbf{q} \in \mathbb{R}^{(D=6)}$, $\mathbf{v} \in \mathbb{R}^{(N=8)}$, $\mathbf{N}_i$ draws entries from the zero-mean unit-variance normal distribution $\mathcal{N}(0,1)$, and $r \in [R = 500]$. For each realization of $\mathbf{X}_r$, we compute $\mathbf{q}_{\text{opt},r}$ by solving Lp-PCA exactly and $\mathbf{q}_{\text{bf},r}$ by means of the proposed Lp-BF algorithm. Then, we compute the Performance Degradation Ratio (PDR) $\Delta_r = 1 - \|\mathbf{X}_r^\top \mathbf{q}_{\text{bf},r}\|_p^p \|\mathbf{X}_r^\top \mathbf{q}_{\text{opt},r}\|_p^{-p}$. In Fig. 3, we illustrate the empirical Cumulative Distribution Function (CDF) of $\{\Delta_r\}_{r \in [R]}$ for $p \in \{0.25, 0.50, 0.75\}$. We observe that all curves exhibit high performance. The curve corresponding to $p = 0.25$ exhibits PDR less than 0.12 with probability 1. The curves corresponding to $p = 0.50$ and 0.75, both attain PDR less than 0.22 with probability 1.

Next, we consider a matrix reconstruction study. That is, we consider $\mathbf{X}_r = \mathbf{X}_{\text{nom}} + \mathbf{N}_r + \mathbf{O}_r \in \mathbb{R}^{(D=8) \times (N=10)}$, where $\mathbf{X}_{\text{nom}} = \sqrt{N\sigma_s^2}\mathbf{Q}\mathbf{V}^\top$, $\mathbf{Q} \in \mathbb{S}_{D \times K}$, $\mathbf{V} \in \mathbb{R}^{N \times K}$, $\mathbf{N}_r$ draws entries from $\mathcal{N}(0, \sigma_n^2)$, $\mathbf{O}_r$ has a single non-zero column at an arbitrary location drawn from $\mathcal{N}(\mathbf{0}_D, \sigma_o^2 \mathbf{I}_D)$, and $r \in [R = 500]$. Above, $\mathbf{X}_{\text{nom}}$ models the rank-K signal-of-interest content carried by $\mathbf{X}_r$, $\mathbf{N}_r$ models Additive White Gaussian Noise (AWGN), $\mathbf{O}_r$ models a high-magnitude/peripheral outlier, and r denotes the r-th realization of noise/outlier. Moreover, the realizations of the noise and outlier signals are statistically independent across realizations. For every $r \in [R]$, we process $\mathbf{X}_r$ by means of PCA (SVD) and the proposed Lp-BF algorithm in Fig. 2, to obtain $\mathbf{Q}_{\text{svd},r} \in \mathbb{S}_{D \times K}$ and $\mathbf{Q}_{\text{bf},r} \in \mathbb{S}_{D \times K}$, respectively. Then, we obtain estimates $\widehat{\mathbf{X}}_{\text{svd},r} = \mathbf{Q}_{\text{svd},r}\mathbf{Q}_{\text{svd},r}^\top \mathbf{X}_r$ and $\widehat{\mathbf{X}}_{\text{bf},r} = \mathbf{Q}_{\text{bf},r}\mathbf{Q}_{\text{bf},r}^\top \mathbf{X}_r$. Then, for each estimate $\widehat{\mathbf{X}}_r \in \{\widehat{\mathbf{X}}_{\text{svd},r}, \widehat{\mathbf{X}}_{\text{bf},r}\}$ we measure the Normalized Reconstruction Error (NRE) $e(\widehat{\mathbf{X}}_r, \mathbf{X}_{\text{nom}}) = \|\widehat{\mathbf{X}}_r - \mathbf{X}_{\text{nom}}\|_F^2 \|\mathbf{X}_{\text{nom}}\|_F^{-2}$. In Fig. 4, we plot the Mean NRE MNRE $= \frac{1}{R}\sum_{r \in [R]} e(\widehat{\mathbf{X}}_r, \mathbf{X}_{\text{nom}})$, for $(\sigma_s^2, \sigma_n^2, K) = (10\text{dB}, (DN)^{-1}, 1)$, $p \in \{0.25, 0.75\}$, and σ_o^2 varying in $\{0, 4, 8, 12\}$dB. We observe that for low values of corruption variance ($\sigma_o^2 \leq 4$dB), all methods exhibit almost identical reconstruction performance. For $\sigma_o^2 = 8$dB, the curves obtained by the proposed Lp-BF algorithm exhibit higher performance than the standard PCA curve. Finally, for $\sigma_o^2 = 12$dB the curve corresponding to PCA, implemented by means of SVD, is clearly misled due to the outlier. In contrast, the Lp-PCA curves exhibit high reconstruction performance compared to standard PCA. In Fig. 5, we plot the MNRE for $(\sigma_s^2, \sigma_n^2, K) = (12\text{dB}, (DN)^{-1}, 2)$, $p \in \{0.25, 0.75\}$, and σ_o^2 varying in $\{0, 4, 8, 12, 16\}$dB. Similar conclusion as in Fig. 4 are drawn. We conclude our studies with a medical-

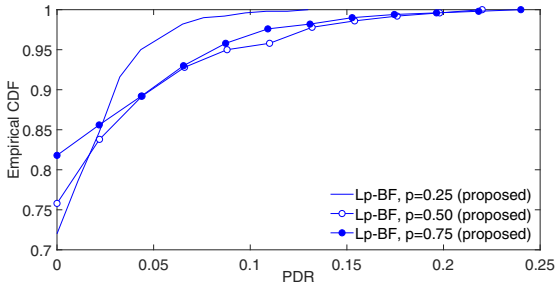

Fig. 3. CDF of PDR with respect to the exact Lp-PCA solution.

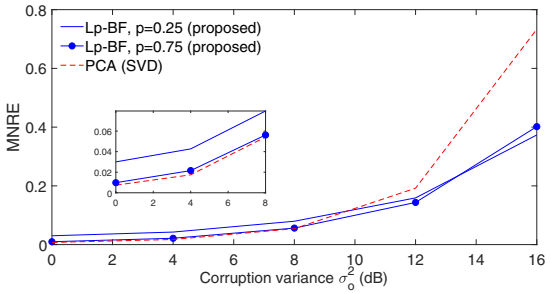

Fig. 5. MNRE vs corruption variance. $(\sigma_s^2, \sigma_n^2, K)$=(12dB, $(DN)^{-1}$, 2).

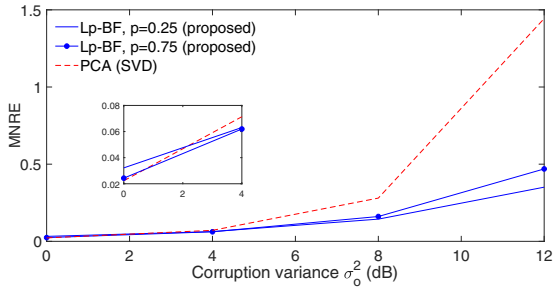

Fig. 4. MNRE vs corruption variance. $(\sigma_s^2, \sigma_n^2, K)$=(10dB, $(DN)^{-1}$, 1).

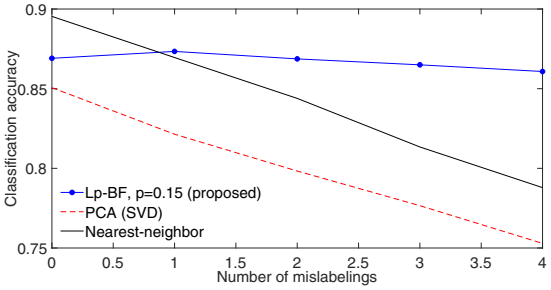

Fig. 6. Classification accuracy vs number of mislabelings.

data classification study. We work with the Breast Cancer Wisconsin (Diagnostic) dataset which includes 569 samples in the form of length $D = 30$ feature-vectors computed from a digitized image of a fine needle aspirate (FNA) of a breast mass [40, 41]. Each sample is labeled either as malignant or benign tissue. There exist 212 malignant and 357 benign tissue-samples in the dataset. We consider that $N_{\text{train}} = 30$ points from each class are available for training and $N_{\text{test}} = 60$ samples are available for testing. We let $\mathbf{X}_b \in \mathbb{R}^{30\times 30}$ and $\mathbf{X}_m \in \mathbb{R}^{30\times 30}$ denote the available benign and malignant training data samples, respectively. Similarly, we let $\mathbf{Y}_b \in \mathbb{R}^{30\times 60}$ and $\mathbf{Y}_m \in \mathbb{R}^{30\times 60}$ denote the available benign and malignant testing data samples, respectively. Importantly, we consider that there is no overlap between training and testing samples. We construct a classifier as follows. During training, we compute $\mathbf{q}_b = \texttt{Lp-BF}(\mathbf{X}_b, p)$ and $\mathbf{q}_m = \texttt{Lp-BF}(\mathbf{X}_m, p)$. Each (unknown) testing sample $\mathbf{y}$ is classified as benign if $(\mathbf{q}_m^\top \mathbf{y})^2 \|\mathbf{y}\|^{-2} < (\mathbf{q}_b^\top \mathbf{y})^2 \|\mathbf{y}\|^{-2}$ and malignant otherwise. To make the classification experiment more challenging, we assume that m benign samples have wrongly been labeled as malignant and m malignant samples have been labeled as benign. We fix $p = 0.15$, let m vary in $\{0, 1, 2, 3, 4\}$, and compute the average classification accuracy, computed over 500 distinct realizations of training/testing data splits and mislabelings. As benchmarks, we include the classification accuracy of PCA[1] and $(k = 1)$-nearest-neighbor classifiers. The average classification accuracy of the three classifiers is reported in Fig. 6. We observe that in the absence of mislabelings $(m = 0)$, the nearest-neighbor classifier exhibits the highest performance followed by the proposed Lp-BF $(p = 0.15)$ classifier with a small difference (about 0.025) in accuracy. For any value of $m \geq 1$, Lp-BF $(p = 0.15)$ outperforms all counterparts, remaining almost unaffected by the mislabelings. In contrast, the performances of the PCA and nearest-neighbor classifiers drops sharply as m increases. For instance, when $m = 4$, Lp-BF $(p = 0.15)$ attains an accuracy of about 0.87 while the PCA and nearest-neighbor classifiers attain performances of about 0.75 and 0.78, respectively.

[1]PCA-based classifier: $\mathbf{q}_b = \texttt{dsv}(\mathbf{X}_b)$ and $\mathbf{q}_m = \texttt{dsv}(\mathbf{X}_m)$. $\texttt{dsv}(\cdot)$ returns the dominant singular vector of its input matrix argument.

V. Conclusions

We presented, for the first time, a converging algorithm for approximating the Lp-quasi-norm $(p \leq 1)$ principal component of a matrix. Then, we followed a subspace deflation approach and presented a novel algorithm for extracting $k \geq 1$ principal components. Numerical studies on synthetic/real data corroborate the outlier resistance of Lp-quasi-norm PCA.

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. Candés, 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] A. Kyrillidis, "Simple and practical algorithms for l_p-norm low-rank approximation," in *Proc. Conf. Uncertainty Artif. Intell.*, Monterey, CA, Aug. 2018, pp. 414-424.

[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, pp. 1-5.

[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] Y. Liu, K. Tountas, D. A. Pados, S. N. Batalama, and M. J. Medley, "L1-Subspace Tracking for Streaming Data," *J. Pattern Recogn.*, vol. 97, pp. 106992:1-106992:13, 2020.

[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,

[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, Jan. 2018.

[24] D. G. Chachlakis and P. P. Markopoulos, "Robust decomposition of 3-way tensors based on L1-norm," in *Proc. SPIE Defense Comm. Sens.*, Orlando, FL, Apr. 2018, pp. 1065807:1–1065807:15.
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 890O:1-109 890O:13.

[28] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm higher-order orthogonal iterations for robust tensor analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Barcelona, Spain, May 2020, pp. 4826-4830.

[29] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm Tucker Tensor Decomposition," in *IEEE Access*, vol. 7, pp. 178454-178465, 2019.

[30] 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, pp. 1658-1661.

[31] 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, pp. 1-5.

[32] D. G. Chachlakis and P. P. Markopoulos, "Combinatorial search for the Lp-norm principal component of a matrix," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2019, pp. 1611-1615.

[33] C. C. Pugh, *Real mathematical analysis*, vol. 11, New York/Heidelberg/Berlin: Springer, 2002.

[34] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, "The Gelfand widths of Lp-balls for $0 < p \leq 1$," in J. Complexity, vol. 26 no. 6, pp. 629-640, 2010.

[35] N. Kwak, "Principal component analysis by L_p-norm maximization," *IEEE Trans. Cybernetics*, vol. 44, no.5, pp. 594-609, 2013.

[36] S. J. Wright, *Primal-dual interior-point methods*, SIAM, 1997.

[37] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*, Reading, MA: Addison-wesley, 1984.

[38] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx, Sep. 2013.

[39] 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.

[40] 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.

[41] 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.