Neurocomputing 000 (2018) 1–9

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust principal component analysis via optimal mean by joint $\ell_{2,1}$ and Schatten p-norms minimization

Xiaoshuang Shi [a], Feiping Nie [b], Zhihui Lai [c], Zhenhua Guo [a,*]

[a] Graduate School at Shenzhen, Tsinghua University, Shenzhen, PR China
[b] School of Computer Science, Northwestern Polytechnical University, Xi'an, Shanxi, PR China
[c] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, PR China

ARTICLE INFO

Article history:
Received 28 November 2016
Revised 4 December 2017
Accepted 21 December 2017
Available online xxx

Communicated by Prof. Zidong Wang

Keywords:
Robust PCA
Optimal mean
$\ell_{2,1}$-norm
Schatten p-norm

ABSTRACT

Since principal component analysis (PCA) is sensitive to corrupted variables or observations that affect its performance and applicability in real scenarios, some convex robust PCA methods have been developed to enhance the robustness of PCA. However, most of them neglect the optimal mean calculation problem. They center the data with the mean calculated by the ℓ_2-norm, which is incorrect because the ℓ_1-norm objective function is used in the following steps. In this paper, we consider a novel robust PCA method that can pursue and remove outliers, exactly recover a low-rank matrix and calculate the optimal mean. Specifically, we propose an optimization model constituted by a $\ell_{2,1}$-norm based loss function and a Schatten p-norm regularization term. The $\ell_{2,1}$-norm used in loss function aims to pursue and remove outliers, the Schatten p-norm can suppress the singular values of reconstructed data at smaller p $(0 < p < 1)$, so it is a better approximation to the rank than the trace norm. Experimental results on benchmark databases demonstrate the effectiveness of our proposed algorithm.

1. Introduction

Finding and exploiting low-dimensional structure in high-dimensional data is a key research topic in many fields, such as image processing, web relevancy data analysis and search, bioinformatics, and so on [1–6]. The original data usually lie in thousands or even millions of dimensions [7,8] in such applications. To alleviate the *curse of dimensionality*, researchers utilize the fact that the data approximately have low intrinsic dimensionality inferred from the low-dimensional subspace [9,10], sparse basis [11–13] or low-dimensional manifold [14–18]. One of the most widely used techniques for dimensionality reduction is principal component analysis (PCA) [19], which finds a lower-dimensional approximating subspace using singular value decomposition (SVD) to form a low-rank approximation. However, standard PCA is brittle with grossly corrupted variables or observations, thus numerous robust PCA methods [20–28] have been proposed to enhance the robustness of PCA.

Recently, several convex robust PCA methods with encouraging results have been developed to improve the robustness of PCA, like robust PCA via proximal gradient with continuation (RPPGC) [9] and robust PCA via outlier pursuit (RPOP) [28]. More precisely, given a low-rank matrix $\mathbf{D} \in \mathbb{R}^{n\times m}$, where n is the number of observations, and m is the number of variables. Suppose:

$$\mathbf{D} = \mathbf{A}_0 + \mathbf{E}_0, \tag{1}$$

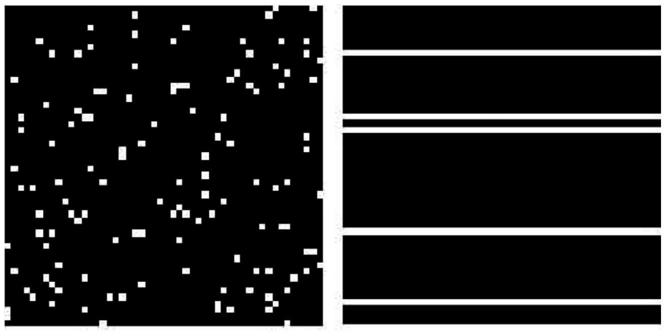
where $\mathbf{A}_0 \in \mathbb{R}^{n\times m}$ is a low-rank matrix and $\mathbf{E}_0 \in \mathbb{R}^{n\times m}$ is a noise matrix. Robust PCA methods [9,28,29] seek the best rank-k estimation of $\mathbf{A}_0$ by solving:

$$min\|\mathbf{D} - \mathbf{A}\|_{\rho}, \quad s.t.\ rank(\mathbf{A}) \le k, \tag{2}$$

where $\|\cdot\|_{\rho}$ represents different norms and indicates certain strategy for enchaining the robustness of PCA. As the ℓ_1-norm is used in Eq. (2), like RPPGC, $\mathbf{E}_0$ should be a matrix with sparsely distributed noise elements (as shown in Fig. 1a), because the ℓ_1-norm minimization, as the minimum convex function to the ℓ_0-norm minimization, can lead to some zero elements in column vectors [9]; as selecting the $\ell_{2,1}$-norm in Eq. (2), like RPOP, $\mathbf{E}_0$ is non-zero in only a fraction of rows (as shown in Fig. 1b), because the $\ell_{2,1}$-norm minimization, as the closest convex function of the $\ell_{2,0}$-norm minimization, encourages the row-sparsity so that it can pursue and remove the outliers. $rank(\mathbf{A})$ is used for noisy data recovery. Since the rank minimization is a non-convex problem, $rank(\mathbf{A})$ is usually replaced by the trace norm (nuclear norm) term $\|\mathbf{A}\|_{\Sigma}$, the minimum convex hull of $rank(\mathbf{A})$. Furthermore, the optimization model composing of the trace norm term does not consider the case that the singular value of reconstructed data could be greatly suppressed, thus Schatten p-norm term suppressing the singular

* Corresponding author.
E-mail address: zhenhua.guo@sz.tsinghua.edu.cn (Z. Guo).

https://doi.org/10.1016/j.neucom.2017.12.034
0925-2312/



(a) (b)

Fig. 1. Two typical types of noise (white points represent noise). (a) Randomly corruption, this case can be handled by the ℓ_1-norm and (b) corrupted data points, which can be better handled by the $\ell_{2,1}$-norm.

values at smaller p $(0 < p < 1)$, which is more approximate to the rank, has been proposed to replace the trace norm term, e.g. minimal shrinkage for noisy data recovery (MSNDR) [29].

Although above mentioned robust PCA methods [9,28,29] can enhance the robustness of PCA in many cases, they neglect the problem of optimal mean calculation. In effect, these robust PCA methods are based on one assumption that the mean of the data is zero. However, in most cases the mean of the data is not zero, so PCA is the best to approximate the given data matrix when the optimal mean is removed [30]. Thus the problem of PCA should be:

$$\min_{\mathbf{b},rank(\mathbf{A})=k} \left\|\mathbf{D} - \mathbf{1}\mathbf{b}^T - \mathbf{A}\right\|_{\rho}, \tag{3}$$

where $\mathbf{b} \in \mathbb{R}^m$ is a mean vector, $\mathbf{1} \in \mathbb{R}^n$ is a column vector with all elements being one.

Besides, the ℓ_2-norm distance based mean in standard PCA is not the correct mean in robust PCA anymore. Therefore, as considering the optimal mean, the reconstruction error in above mentioned robust PCA methods might be reduced.

Motivated by above findings, in this paper, we propose a novel robust PCA method which can pursue and remove outliers, greatly suppress the singular values of reconstructed data, and calculate the optimal mean. Specifically, we present a novel optimization model constituted by a $\ell_{2,1}$-norm based loss function and a Schatten p-norm regularization. In addition, we utilize the popular method, augmented Lagrange multiplier (ALM) [31], to solve the proposed optimization problem, and analyze the time complexity of the optimization method. Finally, experiments on benchmark databases demonstrate the effectiveness of our proposed method.

The rest of this paper is structured as follows. Section 2 introduces the notations and definitions used in this paper. Section 3 presents the proposed robust PCA method including the optimization model, optimization process, and time complexity analysis. Section 4 reports and discusses experimental results on benchmark databases. Finally, Section 5 concludes our work and points out future research work.

2. Notations and definitions

In this paper, matrices are written as bold uppercase letters and vectors are written as bold lowercase letters. For a data matrix $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n\times m}$, its i-th row and j-th column are denoted by $\mathbf{x}^i$ and $\mathbf{x}_j$, respectively. The ℓ_p-norm of the vector $\mathbf{v} \in \mathbb{R}^m$ is defined as $\|\mathbf{v}\|_p = (\sum_{i=1}^m |v_i|^p)^{\frac{1}{p}}$. The ℓ_0-norm of the vector $\mathbf{v}$ is defined as $\|\mathbf{v}\|_0 = \sum_{i=1}^m |v_i|^0$. If $v_i = 0$, $|v_i|^0 = 0$; otherwise, $|v_i|^0 = 1$. The Frobenius norm of the matrix $\mathbf{X}$ is defined as:

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2} = \sqrt{\sum_{i=1}^n \left\|\mathbf{x}^i\right\|_2^2}. \tag{4}$$

The $\ell_{2,1}$-norm of a matrix is widely employed to encourage row-sparsity [32]. It is defined as:

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m x_{ij}^2} = \sum_{i=1}^n \left\|\mathbf{x}^i\right\|_2. \tag{5}$$

For the sake of consistency, the quasi-norm $\ell_{2,0}$-norm of the matrix $\mathbf{X}$ is defined as the number of non-zero rows of $\mathbf{X}$.

The extended Schatten p-norm $(0 < p < \infty)$ [33] of the matrix $\mathbf{X}$ is defined as:

$$\|\mathbf{X}\|_{S_p} = \left(\sum_{i=1}^{min(m,n)} \sigma_i^p\right)^{\frac{1}{p}} = \left(Tr((\mathbf{X}^T\mathbf{X})^{\frac{p}{2}})\right)^{\frac{1}{p}}, \tag{6}$$

where σ_i is the i-th singular value of $\mathbf{X}$. A widely used Schatten norm is the Schatten 1-norm that is defined as:

$$\|\mathbf{X}\|_{S_1} = \sum_{i=1}^{min(m,n)} \sigma_i = Tr((\mathbf{X}^T\mathbf{X})^{\frac{1}{2}}), \tag{7}$$

which is also named as trace norm or nuclear norm, and denoted by $\|\mathbf{X}\|_*$ or $\|\mathbf{X}\|_\Sigma$ in the literature. Similarly, the Schatten 0-norm of the matrix $\mathbf{X}$ is defined as:

$$\|\mathbf{X}\|_{S_0} = \sum_{i=1}^{min(m,n)} \sigma_i^0. \tag{8}$$

If $\sigma_i = 0$, $\sigma_i^0 = 0$; otherwise, $\sigma_i^0 = 1$. Based on this definition, $\|\mathbf{X}\|_{S_0} = rank(\mathbf{X})$.

3. Robust PCA via optimal mean (RPOM)

In this section, we present the novel proposed robust PCA method, including an optimization model, the optimization process and computation complexity analysis.

3.1. Optimization model

Suppose that the observed data matrix $\mathbf{D} \in \mathbb{R}^{n\times m}$ includes γn randomly located outliers (corrupted points) and $(1-\gamma)n$ clean points. Then let the matrix $\mathbf{A}_0 \in \mathbb{R}^{n\times m}$ represent n clean points by subtracting a mean vector $\mathbf{b}_0 \in \mathbb{R}^m$, and the noise matrix $\mathbf{E}_0 \in \mathbb{R}^{n\times m}$ have γn non-zero rows, so $\mathbf{D} = \mathbf{A}_0 + \mathbf{E}_0 + \mathbf{1}\mathbf{b}_0^T$, where $\mathbf{1} \in \mathbb{R}^n$ is a column vector with all elements being one and $\mathbf{b}_0$ is also a column vector. The problem of robust PCA via optimal mean can then be formulated as follows:

Problem 1. Given $\mathbf{D} = \mathbf{A} + \mathbf{E} + \mathbf{1}\mathbf{b}^T$, where $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{E}$ are unknown, $\mathbf{A}$ is known to be a low rank matrix, $\mathbf{b}$ is a mean vector and $\mathbf{E}$ is known to be a matrix with some non-zero rows, we would like to recover $\mathbf{A}$ and $\mathbf{b}$.

We can reformulate the Problem 1 as the following optimization problem:

$$\min_{\mathbf{A},\mathbf{b}} \left\|\mathbf{D} - \mathbf{A} - \mathbf{1}\mathbf{b}^T\right\|_{2,0} + \lambda rank(\mathbf{A}). \tag{9}$$

However, both the $\ell_{2,0}$-norm minimization and the rank minimization are NP-hard problems which are hard to be solved efficiently [9,34]. In order to obtain a tractable optimization problem, Eq. (9) is relaxed by replacing the $\ell_{2,0}$-norm with the $\ell_{2,1}$-norm, and the rank with the Schatten p-norm. The $\ell_{2,1}$-norm minimization is the closest convex function of the $\ell_{2,0}$-norm minimization, so if $\mathbf{D} - \mathbf{A} - \mathbf{1}\mathbf{b}^T$ is sufficiently row-sparse, minimizing