

An ADMM Approach for Constructing Abnormal Subspace of Sparse PCA

Wenmin Jiang[1], Ying Zhao[1(✉)], and Kazi Abir Adnan[2]

[1] Tsinghua University, Beijing, China
jwm17@mails.tsinghua.edu.cn, yingz@tsinghua.edu.cn
[2] University of Melbourne, Melbourne, Australia
kazi.adnan@unimelb.edu.au

Abstract. Despite the popularity of principal component analysis (PCA) as an anomaly detection technique, the main shortage of PCA-based anomaly detection models is their interpretability. Constructing the abnormal subspace of PCA (i.e., the subspace spanned by the least significant principal components (PCs)), with sparse and orthogonal loading vectors provides a means of anomaly interpretation. However, solving all abnormal sparse PCs one by one through semi-definite programming is time consuming. In this paper, we derive an adapted projection deflation method for extracting least significant PCs and propose an alternating direction method of multipliers (ADMM) solution for constructing the sparse abnormal subspace. Our experiments on two real world datasets showed that the proposed ADMM solution achieved comparable detection accuracy and sparsity as the SDP solution and is 10 times more efficient, which makes it more suitable for application domains with higher dimensions.

Keywords: Anomaly detection · Sparse PCA · ADMM

1 Introduction

Principal Component Analysis (PCA) is one of the best-known statistical analysis techniques for detecting anomalies and has been applied to many kinds of data, such as network intrusion detection, failure detection in production systems, and so on [1]. Although PCA and derived techniques are not the only solutions for unsupervised anomalies detection, they are among the most widely used ones. However, traditional PCA-based anomaly detection models are not suitable for anomaly interpretation (i.e., pinpointing the causes of detected anomalies) [2] and are treated as black-box techniques. Recently, detection models based on sparse PCA [3,4] were proposed to offer better interpretation since sparse PCA can improve the interpretability of PCA's dimensionality-reduced subspace.

This work is funded by the National Key Research and Development Program of China (Grant No. 2016YFB0200100).

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12253, pp. 709–717, 2020.
https://doi.org/10.1007/978-3-030-58814-4_58

We have investigated the anomaly interpretation problem of PCA-based models in our previous publication [4]. In [4], we proposed an abnormal subspace sparse PCA (ASPCA) model that can find the abnormal subspace defined by sparse and orthogonal principal components (PCs). The proposed ASPCA model prioritizes the sparsity of the abnormal subspace by extracting PCs with least significant variances through a semi-definite programming (SDP) approach. We project each of the data record onto the abnormal subspace and calculate their projection lengths. For those detected anomalies that have higher projection lengths, the ASPCA model interprets its detection results by identifying the set of projections that contribute the most. Although this ASPCA model can provide desirable detection and interpretation to individual anomalies, it suffers from high computational costs due to the SDP optimization. In this paper, we adapt a projection deflation method to extract least significant principal components, and derive an alternating direction method of multipliers (ADMM) solution to improve the efficiency. We make the following contributions:

- Derive an ADMM solution for backward ASPCA that guarantees the orthogonality of extracted principal components through an adapted projection deflation method.
- Conduct experiments involving two real world datasets to show that the proposed ADMM solution can achieve comparable detection accuracy and sparsity as the SDP solution and is 10 times more efficient.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 introduces the ASPCA model and the proposed adapted projection deflation method. Section 4 describes an ADMM solution. Section 5 presents an experimental evaluation on the efficiency of the ADMM solution. Finally, Sect. 6 provides some concluding remarks.

2 Related Works

PCA is mostly known as a dimension reduction tool [5], but it is also widely used as an anomaly detection method thanks to its scalability and communication efficiency [1]. Recently, the sparse PCA introduced by Jolliffe *et al.* [6] showed plausible performance for anomaly detection. Luo *et al.* [7] used sparse PCA to monitor industrial processes and showed that the sparse PCA-based method had better monitoring performance than the PCA-based methods. Yu *et al.* [8] proposed a nonlinear and sparse PCA method (RNSPCA) for industrial as well. Sparse PCA also improves the interpretability of PCA's dimensionality-reduced subspace, which can be used to interpret detected anomalies. Jiang *et al.* [3] proposed a joint sparse PCA (JSPCA) model to achieve a sparse representation of the abnormal subspace and identify the set of features that distinguish anomalies. Bin *et al.* [4] proposed the ASPCA model to construct the abnormal principal components (PCs) with sparse and orthogonal loading vectors and interpret detected anomalies by identifying the set of such PCs on which they have large projection values.

Various methods solving the sparse PCA problem were proposed in the literature, for example [9] and [10]. Aspremont *et al.* [10] proposed a SDP relaxation to the sparse PCA optimization problem. Recently, the alternating direction method of multipliers [11] was used for solving the sparse PCA problem more efficiently [12] and extended by Vu *et al.* [13] by forming a novel convex relaxation of sparse principal subspace estimation based on the convex hull of rank-d projection matrices (i.e., Fantope). In this paper, we examine these techniques to improve the efficiency of the backward ASPCA model.

3 Sparse PCA for Anomaly Detection and Interpretation

We first introduce the notations used in this paper. Bold uppercase letters such as $\mathbf{X}$ denote a matrix and $\mathbf{X}_{ij}$ is the entry of $\mathbf{X}[i,j]$. Bold lowercase letters such as $\mathbf{v}$ denote a column vector. Greek letters such as λ, β are coefficients. $||\mathbf{X}||_{1,1}$ is the $L_{1,1}$ norm of $\mathbf{X}$ as $||\mathbf{X}||_{1,1} = \mathbf{1}^T|\mathbf{X}|\mathbf{1}$. $Tr(\mathbf{X})$ represents the trace of matrix $\mathbf{X}$. $<\mathbf{X},\mathbf{Y}>$ represents the matrix inner product defined as $<\mathbf{X},\mathbf{Y}> = Tr(\mathbf{X}^T Y)$. $Card(\mathbf{X})$ denotes the cardinality (number of non-zero elements) of matrix $\mathbf{X}$. $\mathbf{I}$ is the identity matrix. $\mathbb{S}_+^p$ is the set of all symmetric positive semi-definite matrices in $\mathbb{R}^{p\times p}$, where p is the number of dimensions.

Principal Component Analysis (PCA) captures the principal components of a multi-dimensional dataset defined by a set of orthogonal eigenvectors with the highest variance. Given a p-dimensional dataset, an anomaly detection model can be constructed by forming a normal subspace (defined by the first k principal components returned by PCA) and an abnormal subspace (defined by the remaining d principal components).

In [4], we proposed an interpretable PCA-based anomaly detection model, the Abnormal Subspace Sparse PCA (ASPCA) model, which can detect and interpret anomalies using the abnormal subspace directly. In particular, the ASPCA model tries to find the abnormal subspace defined by sparse and orthogonal principal components. Given a $n\times p$ data matrix $\mathbf{D}$, its covariance matrix $\mathbf{A} = \mathbf{D}^T\mathbf{D}$, and a sparsity constraint constant c, for each $i = p, ..., p-d+1$, the backward ASPCA (ASPCA-B) model tries to solve:

$$\begin{aligned} &\underset{\mathbf{v}_i}{\text{argmin}}\ \mathbf{v}_i^T\mathbf{A}\mathbf{v}_i \\ &s.t.\ \mathbf{v}_i^T\mathbf{v}_i = 1,\ \mathbf{v}_i^T\mathbf{v}_j = 0\ \forall i < j \le p, Card(\mathbf{v}_i) \le c. \end{aligned} \tag{1}$$

When we extract d loading vectors $\mathbf{v}_{p-d+1}, ..., \mathbf{v}_p$ using Eq. 1 to span a subspace $\mathbf{V}$, we make sure that the orthogonal complement of $\mathbf{V}$ has major variance for describing the normal patterns in the dataset. Given a p-dimensional data record $\mathbf{y}$, the the squared prediction error (SPE) can be calculated as $||\hat{\mathbf{y}}||_2^2 = \hat{\mathbf{y}}^T\hat{\mathbf{y}} = \sum_{i=p-d+1}^{p}(\mathbf{v}_i^T\mathbf{y})^2$ for judging whether $\mathbf{y}$ is an anomaly or not. The loading vectors that contribute the most to the SPE score can be used to interpret the cause of the anomaly since they are sparse and understandable to users.

Instead of solving Eq. 1 with the orthogonality constraint, we can also adopt a deflation method [14] that ensures orthogonality to extract these components sequentially, in which case we drop the orthogonality constraint from Eq. 1 and the objective function of our ASPCA-B model becomes:

$$\begin{aligned} &\underset{\mathbf{v}}{\text{argmin}} \ \mathbf{v}^T\mathbf{A}\mathbf{v} \\ &s.t.\ \mathbf{v}^T\mathbf{v} = 1, Card(\mathbf{v}) \leq c. \end{aligned} \tag{2}$$

Here, we adapt the original deflation method proposed in [14] and make it usable for extracting the least significant components. By the symmetric Schur decomposition, we have $\mathbf{I} - \mathbf{v}_i\mathbf{v}_i^T = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix with $\mathbf{\Lambda}_{ii} = 0$ and $\mathbf{\Lambda}_{jj} = 1$ for $j \neq i, 1 \leq j \leq p$. By eliminating the ith column of $\mathbf{Q}$ we obtain a $p \times (p-1)$ matrix $\hat{\mathbf{Q}}$. Now we can use the deflation as $\hat{\mathbf{A}} = \hat{\mathbf{Q}}^T\mathbf{A}\hat{\mathbf{Q}}$, and the following proposition states its correctness.

Proposition 1. *If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of $\mathbf{A} \in \mathbb{S}_+^p$, $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ are the corresponding eigenvectors, and $\hat{\mathbf{A}} = \hat{\mathbf{Q}}^T\mathbf{A}\hat{\mathbf{Q}}$ for some $i \in 1, \cdots, p$, then $\hat{\mathbf{A}}$ has eigenvectors $\hat{\mathbf{v}}_1, \cdots, \hat{\mathbf{v}}_{i-1}, \hat{\mathbf{v}}_{i+1}, \cdots, \hat{\mathbf{v}}_p$ with corresponding eigenvalues $\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_p$, where $\hat{\mathbf{v}}_j = \hat{\mathbf{Q}}^T\mathbf{v}_j$, for $j \neq i, 1 \leq j \leq p$.*

Proof. $\forall j \neq i, \hat{\mathbf{A}}\hat{\mathbf{v}}_j = \hat{\mathbf{Q}}^T\mathbf{A}\hat{\mathbf{Q}}\hat{\mathbf{Q}}^T\mathbf{v}_j = \hat{\mathbf{Q}}^T\mathbf{A}(\mathbf{I} - \mathbf{v}_i\mathbf{v}_i^T)\mathbf{v}_j = \hat{\mathbf{Q}}^T\mathbf{A}\mathbf{v}_j = \hat{\mathbf{Q}}^T\lambda_j\mathbf{v}_j = \lambda_j\hat{\mathbf{v}}_j$.

4 Optimization Methods

4.1 Solving Backward ASPCA with SDP

We transform Eq. 1 without the orthogonality constraint $\mathbf{v}_j^T\mathbf{v}_i = 0, \forall 1 \leq j < i$ to Eq. 3 through a SDP relaxation.

$$\begin{aligned} &\underset{\mathbf{X}_i \in \mathbb{S}^p}{\text{argmin}} \ Tr(\mathbf{A}\mathbf{X}_i) \\ &s.t.\ \mathbf{X}_i \succeq 0, rank(\mathbf{X}_i) = 1,\ Tr(\mathbf{X}_i) = 1, Card(\mathbf{X}_i) < k^2, \end{aligned} \tag{3}$$

where $\mathbf{X}_i$ is a positive semi-definitive matrix with the constraint $rank(\mathbf{X}_i) = 1$, which can be uniquely decomposed as $\mathbf{X}_i = \mathbf{v}_i\mathbf{v}_i^T$. With $\mathbf{X}_i = \mathbf{v}_i\mathbf{v}_i^T$, $Tr(\mathbf{X}_i) = 1$ is equivalent to $\mathbf{v}_i^T\mathbf{v}_i = 1$, $Card(\mathbf{X}_i) \leq k^2$ is equivalent to $Card(\mathbf{v}_i) \leq k$, and we have $\mathbf{v}_i^T\mathbf{A}\mathbf{v}_i = Tr(\mathbf{A}(\mathbf{v}_i\mathbf{v}_i^T)) = Tr(\mathbf{A}\mathbf{X}_i)$.

Now let $\mathbf{V}_i = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_i)$ and $\mathbf{R}_i = \mathbf{V}_i\mathbf{V}_i^T$, the orthogonality constraint $\mathbf{v}_j^T\mathbf{v}_i = 0, \forall 1 \leq j < i$ is equivalent to $||\mathbf{V}_{i-1}^T\mathbf{v}_i||_2^2 = Tr(\mathbf{R}_{i-1}\mathbf{X}_i) = 0$. After moving the sparsity constraint into the objective function and dropping the non-convex constraint $rank(\mathbf{X}_i) = 1$, we have an objective function that can be solved by semidefinite programming (SDP) as in Eq. 4.

$$\begin{aligned} &\underset{\mathbf{X}_i \in \mathbb{S}^p}{\text{argmin}} \ Tr(\mathbf{A}\mathbf{X}_i) + \lambda||\mathbf{X}_i||_{1,1} \\ &s.t.\ \mathbf{X}_i \succeq 0,\ Tr(\mathbf{X}_i) = 1, Tr(\mathbf{R}_{i-1}\mathbf{X}_i) = 0. \end{aligned} \tag{4}$$

We refer to this method as **backward SDP ASPCA (SDP-B)**.

4.2 Solving Backward ASPCA with ADMM

Following the SDP relaxation framework, Eq. 2 can be transferred to the following optimization problem,

$$\begin{aligned} &\underset{\mathbf{X}\in\mathbb{S}_+^p}{\text{argmin}} \ Tr(\mathbf{AX}) + \lambda||\mathbf{X}||_{1,1} \\ &s.t.\ \mathbf{X} \succeq 0,\ Tr(\mathbf{X}) = 1. \end{aligned} \tag{5}$$

Let $\mathcal{C}$ denote the simplex of the cone of the semidefinite matrices, i.e., $\mathcal{C} = \{\mathbf{X} \in \mathbb{S}_+^p : \mathbf{X} \succeq 0,\ Tr(\mathbf{X}) = 1\}$, $I_{\mathcal{C}}(\mathbf{X})$ denote the indicator function of set $\mathcal{C}$, i.e., $I_{\mathcal{C}}(\mathbf{X}) = 0$ if $\mathbf{X} \in \mathcal{C}$, and ∞ otherwise. By introducing a new variable $\mathbf{Y}$ and following the augmented Lagrangian framework, we can solve Eq. 5 using the alternating direction method of multipliers (ADMM) as follows,

$$\begin{aligned} &\underset{\mathbf{X}}{\text{argmin}} \ I_{\mathcal{C}}(\mathbf{X}) + Tr(\mathbf{AX}) + \lambda||\mathbf{Y}||_{1,1} \\ &s.t.\ \mathbf{X} - \mathbf{Y} = 0. \end{aligned} \tag{6}$$

The augmented Lagrangian associated with Eq. 6 has the form

$$\begin{aligned} &\mathcal{L}_\rho(\mathbf{X}, \mathbf{Y}; \mathbf{U}) = <\mathbf{A}, \mathbf{X}> + I_{\mathcal{C}}(\mathbf{X}) + \lambda||\mathbf{Y}||_{1,1} \\ &+ \frac{\rho}{2}||\mathbf{X} - \mathbf{Y}||_F^2 - <\mathbf{U}, \mathbf{X} - \mathbf{Y}>, \end{aligned} \tag{7}$$

where $\mathbf{U}$ is the Lagrange multiplier associated with the constraint $\mathbf{X} = \mathbf{Y}$ and $\rho > 0$ is a penalty parameter.

Now we minimize Eq. 7 with respect to $\mathbf{X}$ and $\mathbf{Y}$ alternatively and update $\mathbf{X}^{t+1}$ and $\mathbf{Y}^{t+1}$ iteratively by

$$\begin{aligned} \mathbf{X}^{t+1} &= \mathcal{P}_{\mathcal{C}}(\mathbf{Y}^t - \mathbf{U}^t/\rho - \mathbf{A}/\rho), \\ \mathbf{Y}^{t+1} &= Shrink(\mathbf{X}^{t+1} - \mathbf{U}^t/\rho, \frac{\lambda}{\rho}). \end{aligned} \tag{8}$$

where $\mathcal{P}_{\mathcal{C}}()$ calculates the projection on $\mathcal{C}$ that can be solved in $O(p \log p)$ [12] and $Shrink()$ is the shrinkage operator. Lastly, the Lagrange multiplier $\mathbf{U}^{t+1}$ is updated by $\mathbf{U}^t + (\mathbf{X}^t - \mathbf{Y}^t)/\rho$. We choose to terminate our ADMM when

$$max(||\mathbf{X}^t - \mathbf{Y}^t||_F^2, \rho^2||\mathbf{Y}^t - \mathbf{Y}^{t-1}||_F^2) \leq d\epsilon^2, \tag{9}$$

where ϵ is a convergence parameter. As the resultant $rank(\mathbf{X})$ might not be 1, we use the dominant eigenvector of $\mathbf{X}$ as the approximate solution for $\mathbf{v}$. Finally, we use the proposed adapted projection deflation method to annihilate $\mathbf{v}$ to prepare for the next eigenvector. We refer to this method as **backward ADMM ASPCA (ADMM-B)**, and show its procedure in Algorithm 1.

Algorithm 1. Backward ADMM ASPCA (ADMM-B)

Input: $\mathbf{A}, d, \lambda, \rho, \epsilon$

d is the number of abnormal subspace PCs; λ is the sparsity penalty parameter; ρ is the Lagrangian penalty parameter; ϵ is the convergence parameter.

Output: $\mathbf{V}$

for $i = p$ to $p - d + 1$ **do**
  $\mathbf{V}_{i+1} \leftarrow (\mathbf{v}_{i+1}, \cdots, \mathbf{v}_p)$;
  Schur decomposition $\mathbf{Q}^T \mathbf{\Lambda} \mathbf{Q} = \mathbf{I} - \mathbf{V}_{i+1}\mathbf{V}_{i+1}^T$;
  $\hat{\mathbf{Q}} \leftarrow$ the first i columns of $\mathbf{Q}$;
  $\mathbf{A}_i \leftarrow \hat{\mathbf{Q}}^T \mathbf{A} \hat{\mathbf{Q}}$;
  $\mathbf{Y}_i^0 \leftarrow 0, \mathbf{U}_i^0 \leftarrow 0, t \leftarrow 0$;
  repeat
    $\mathbf{X}_i^{t+1} \leftarrow P_{\mathcal{C}}(\mathbf{Y}_i^t - \mathbf{U}_i^t/\rho - \mathbf{A}_i/\rho)$;
    $\mathbf{Y}_i^{t+1} \leftarrow Shrink(\mathbf{X}_i^{t+1} - \mathbf{U}_i^t/\rho, \frac{\lambda}{\rho})$;
    $\mathbf{U}_i^{t+1} \leftarrow \mathbf{U}_i^t + (\mathbf{X}_i^{t+1} - \mathbf{Y}_i^{t+1})/\rho$;
    $t \leftarrow t + 1$;
  until $max(||\mathbf{X}_i^t - \mathbf{Y}_i^t||_F^2, \rho^2||\mathbf{Y}_i^t - \mathbf{Y}_i^{t-1}||_F^2) \leq d\epsilon^2$
  $\hat{\mathbf{v}}_i \leftarrow$ the dominant eigenvector of $\mathbf{X}_i$;
  $\mathbf{v}_i \leftarrow \hat{\mathbf{Q}}\hat{\mathbf{v}}_i$;
end for
$\mathbf{V} \leftarrow (\mathbf{v}_{p-d+1}, ..., \mathbf{v}_p)$;
return $\mathbf{V}$;

5 Experiments

The various ASPCA models were evaluated on their anomaly detection performance, sparsity of loading vectors of their abnormal PCs, and efficiency using two real-world datasets **Breast-Cancer** and **KDD99**.

Breast Cancer Wisconsin (Diagnostic) Dataset [15] provides features to distinguish malignant and benign tumors. Our dataset includes 357 benign records and 20 malignant records sampled randomly from the original dataset as anomalies. For each data record, there are 30 features describing characteristics of the cell nuclei present in a digitized image of a *fine needle aspirate* (FNA) of a breast mass. All 30 real-valued features were deducted by the mean values and linearly scaled to $[-1, 1]$. **KDD 99 Intrusion Dataset** [16] is widely used for anomaly and intrusion detection. Each instance is a connection record classified as normal or one of 22 classes of attacks. We kept all 97278 normal instances and sampled 5173 abnormal records. We followed a similar preprocessing procedure as in [3]. There are 41 features and all features were deducted by the mean values and linearly scaled to $[-1, 1]$.

We summarize the various evaluated models here:

- **PCA:** A standard PCA is our baseline model for anomaly detection performance and PCs sparsity.
- **SDP-B:** The backward ASPCA model optimized using SDP presented in [4] is another baseline model in our experiments.

- **ADMM-B:** The proposed ADMM solution can efficiently optimize the backward ASPCA objective function.
- **ADMM-A:** Another ADMM-based sparse PCA model presented in [13] was also considered in our comparison, which forms a convex relaxation of sparse principal subspace estimation based on the convex hull of rank-d projection matrices. The ADMM-A model extracts all d abnormal PCs simultaneously, but cannot guarantee the orthogonality of extracted PCs.

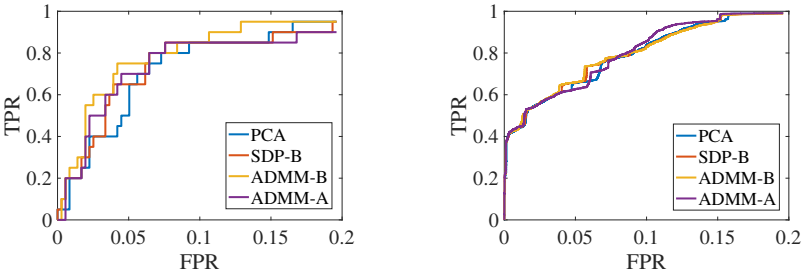
For all the models, the most important parameter is the number of abnormal subspace PCs d. We chose d such that the abnormal subspace PCs explain roughly 5% of the total variance [2] for all the models as shown in Table 1. For all the ASPCA models, we chose the PC sparsity penalty parameter λ as suggested in [4]. For ADMM models, the Lagrange penalty parameter ρ was set to be equal to λ as suggested in [11]. The convergence parameter ϵ was set to be 0.01. Finally, we implemented all methods with MATLAB and CVX, and performed all experiments on a laptop computer with 8 GB memory and a Intel Core i5 1.4 GHz CPU.

Table 1. Parameters for various models.

	d (PCA)	d (ASPCA models)	λ (sparsity penalty)
Breast-Cancer	19	15	0.5
KDD99	34	34	100

Anomaly Detection. We used the abnormal subspace PCs obtained by various models to predict all the anomalies in the dataset and generated ROC curves to evaluate the anomaly detection performance. The results are shown in Fig. 1. Recall that we chose d to make sure that the abnormal subspace PCs obtained by various models explain roughly 5% of the total variance. Under this condition, we can see that the various ASPCA models achieved similar anomaly detection performance as the standard PCA model indicated by the similarities of their ROC curves for both Breast-Cancer and KDD99 datasets.

Sparsity Evaluation. The next set of experiments were designed to compare the sparsity of the loading matrix generated by various ASPCA models and we used the result of PCA as our baseline. We used three metrics to evaluate the sparsity of the loading matrix of the abnormal PCs, namely, $||V||_{1,1}$, $Card_{0.1}$ (number of entries with absolute values bigger than 0.1), and $Card_{0.01}$ (number of entries with absolute values bigger than 0.01), and showed the results in Table 2. We can see that all ASPCA models improved the sparsity of the loading matrix greatly over the baseline. The ADMM-B model indeed achieved the same sparsity performance as SDP-B. The ADMM-A model achieved a worse sparsity performance, which suggests that it is not suitable for constructing the sparse abnormal subspace for anomaly interpretation.



(a) Breast-Cancer (b) KDD99

Fig. 1. ROC curve of PCs of Breast-Cancer and KDD99 obtained by various models.

Table 2. Sparsity on Breast-Cancer and KDD99.

Dataset	Method	$\|\|V\|\|_{1,1}$	$card_{0.1}$	$card_{0.01}$
Breast-Cancer	PCA	70.12	242	483
	SDP-B	26.77	59	80
	ADMM-B	27.37	52	112
	ADMM-A	49.05	149	329
KDD99	PCA	91.28	237	630
	SDP-B	52.08	95	194
	ADMM-B	52.07	93	192
	ADMM-A	76.03	203	421

Efficiency. We showed the runtime of various models on the two datasets in Table 3. We can see that all ASPCA models have higher time complexity than the standard PCA for computing sparse PCs. As expected, the ADMM-B model indeed improves efficiency over the SDP-B model. The ADMM-B model ran about 40 and 10 times faster than the SDP-B model on Breast-Cancer and KDD99, respectively. The ADMM-A model solves all the abnormal PCs simultaneously and is faster than the ADMM-B model, which solves the abnormal PCs one by one. However, the ADMM-A model cannot achieve the desirable sparsity performance.

Table 3. Runtime (in seconds).

Models	Breast-Cancer	KDD99
PCA	0.001	0.13
SDP-B	13.045	27.56
ADMM-B	0.335	2.60
ADMM-A	0.009	0.21

6 Conclusions

We proposed an ADMM solution to extract sparse abnormal subspace of PCA with sparse and orthogonal PCs for anomaly detection and interpretation. The proposed ADMM solution is much more efficient than a SDP solution, which makes it possible to apply to datasets with higher dimensions. In the further, we will study the efficiency and interpretability of this anomaly detection model on more application domains.

References

1. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 1–58 (2009)
2. Ringberg, H., Soule, A., Rexford, J., Diot, C.: Sensitivity of PCA for traffic anomaly detection. In: The 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, pp. 109–120. ACM (2007)
3. Jiang, R., Fei, H., Huan, J.: A family of joint sparse PCA algorithms for anomaly localization in network data streams. IEEE Trans. Knowl. Data Eng. **25**(11), 2421–2433 (2013)
4. Bin, X., Zhao, Y., Shen, B.: Abnormal subspace sparse PCA for anomaly detection and intepretation. In: The ACM SIGKDD Workshop on Outlier Detection and Description, ODD 2015, pp. 1–10 (2015)
5. Jolliffe, I.: Principal Component Analysis. Wiley Online Library, Hoboken (2002)
6. Jolliffe, I.T., Trendafilov, N.T., Uddin, M.: A modified principal component technique based on the LASSO. J. Comput. Graph. Stat. **12**(3), 531–547 (2003)
7. Luo, L., Bao, S., Mao, J., Tang, D.: Fault detection and diagnosis based on sparse PCA and two-level contribution plots. Ind. Eng. Chem. Res. **56**(1), 225–240 (2017)
8. Yu, H., Khan, F., Garaniya, V.: A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes. AIChE J. **62**(5), 1494–1513 (2016)
9. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. **15**(2), 265–286 (2006)
10. d'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse PCA using semidefinite programming. SIAM Rev. **49**(3), 434–448 (2007)
11. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2011)
12. Ma, S.: Alternating direction method for multipliers for sparse principal component analysis. J. Oper. Res. Soc. China **1**, 253–274 (2013)
13. Vu, V.Q., Cho, J., Lei, J., Rohe, K.: Fantope projection and selection: a near-optimal convex relaxation of sparse PCA. In: Advances in Neural Information Processing Systems, vol. 26, pp. 2670–2678 (2013)
14. Mackey, L.: Deflation methods for sparse PCA. In: Advances in Neural Information Processing Systems, vol. 22, pp. 1017–1024 (2009)
15. Breast Cancer Wisconsin (Diagnostic) Data Set (1995)
16. KDD Cup 1999 Data (1999)