Dimensionality-reduced Secure Outlier Detection on Union of Subspaces

Kunzan Liu, Yuchen Jiao
Department of Electronic Engineering
Tsinghua University
Beijing, China
{lkz18,jiaoyc18}@mails.tsinghua.edu.cn

Ye Jin, Xu Xiang
National Key Laboratory on Blind Signal Processing
Chengdu, China
beidou_jy@Hotmail.com
xiangxu1402@163.com

Yuantao Gu
Department of Electronic Engineering
Tsinghua University
Beijing, China
gyt@tsinghua.edu.cn

***Abstract*—In the problem of outlier detection (OD) on a union of subspaces (UoS), inliers are assumed to lie around a union of low-dimensional subspaces, and the goal is to detect the outliers that are not close to any of these subspaces. Among various algorithms, sparse self-representation-based ones have attracted much attention because of their theoretical performance guarantee. However, these algorithms need direct access to all raw data, and thus have poor data security and privacy protection capability. To solve this problem, in this paper we propose a new algorithm called dimensionality-reduced secure outlier detection (DrSOD), which uses random projection as a preprocessing step to avoid direct access to the raw data. We theoretically prove that DrSOD can correctly detect outliers with overwhelming probability under connectivity assumptions. In addition, the random projection step improves the computational efficiency of the algorithm. Experiments on synthetic and real-world datasets also demonstrate the effectiveness and efficiency of DrSOD.**

Keywords-random projection; data security and privacy; outlier detection; self-representation; union of subspaces

I. INTRODUCTION

In many computer vision applications such as motion segmentation [1] and face recognition [2], high-dimensional datasets are frequently discovered to have low-dimensional structure called Union of Subspaces (UoS) [3], where data points are assumed to lie around a union of low-dimensional subspaces. Learning systems based on UoS model has received much attention, such as subspace clustering and subspace learning [4]. However, real-world datasets often involve outliers, which neither belong to any subspace nor exhibit low-dimensional structure. Generally, detecting and removing these outliers before further processing can greatly improve the efficiency and precision of the system. Therefore, outlier detection (OD) on UoS model is a significant issue in practical applications.

A maturing increasingly crucial problem in the implementation of outlier detection is the data security and privacy [5]. However, in many existing detection algorithms, analysts need the access to all raw data, which is harmful to data security [6], [7], [8], [9]. For example, in face recognition tasks, the raw data, i.e., face images, are visible, which may reveal personal privacy. Therefore, detecting outliers without direct access to the raw data is necessary for privacy protection.

There are several classes of OD methods. RANSAC [6] is based on fitting a subspace for points randomly selected, and indicating outliers and inliers according to the estimated subspaces. However, the objective problem is usually nonconvex, and its performance heavily relies on good initializations. Another class of algorithms uses the fact that outliers have lower similarities with other data points compared with inliers [7]. While these algorithms require the inliers to be well distributed and densely sampled within all subspaces, which is hard to be satisfied in practice. Other methods solve convex optimization problems with robustness to outliers [8], [9]. However, when dealing with multiple inlier subspaces, these methods treat the union as a single linear subspace, and have large complexity because of the high dimension. Moreover, none of the above methods take data security and privacy into account because they directly process the raw data.

Recently, You et al [10] have enabled OD under a more general condition by incorporating self-representation, which is a widely used technique and has immensely boosted the performance of many data processing tasks. Empirical study shows that this algorithm achieves the state-of-the-art performance. Moreover, it is theoretically proved that this algorithm, under mild assumptions, can correctly detect all outliers as long as the so-called Subspace Preserving (SP) property is satisfied. However, this algorithm still does not consider data security and privacy. Moreover, the computational cost can be extremely large when the dimension of the data points is high [4].

Motivated by privacy protection and computational efficiency of OD, in this paper we propose a new algorithm named Dimensionality-reduced Secure Outlier Detection (DrSOD). It adopts random projection as a preprocessing step. Since that random projection greatly changes the data representation, it is hard to get private information from the preprocessed data. Fig. 1 illustrates this by comparing the raw data (Fig. 1(a)), which is a human face image, and the preprocessed data (Fig. 1(b)). It is obvious that from Fig. 1(b) one cannot recognize the person. Besides privacy protection, random projection also reduces the dimension of the data, and thus improves the computational efficiency of the algorithm.

978-1-7281-8620-7/21/$31.00 ©2021 IEEE

(a) original image (b) preprocessed image

Fig. 1. An example to show the security and privacy consideration of DrSOD. (a) is the original image and (b) is the same image after preprocessing.

After using random projection, an important question that needs to be answered is that whether the detection performance will be degraded. To this end, we theoretically analyze the correctness of DrSOD. Theoretical results show that when the random matrices satisfy the so-called Johnson-Lindenstrauss (JL) Property, random projection will not greatly harm the performance. JL Property is a very mild condition and is satisfied by numerous types of random matrices. They can preserve the inner product between two data points and the distance between two subspaces after mapping [11]. We also conduct experiments using synthetic and real-world datasets. Satisfactory results illustrate the practicability and efficiency of our algorithm.

The theoretical correctness confirmation is the main contribution of this paper. There are some similar theoretical works like [12] which has studied the perturbation of random projection on various subspace clustering algorithms based on the Subspace Restricted Isometry Property [13]. However, random projection has never been applied in self-representation-based OD algorithms before to our best knowledge. Moreover, analysis in this paper shows that the SP property holds with high probability after random projection. Because SP property is the premise of numerous subspace learning problems, our results indicate that random projection can contribute to more applications.

II. PRELIMINARIES

A. Random projection

Random projection, also known as random compression, projects the raw data $\boldsymbol{x} \in \mathbb{R}^N$ onto a medium-dimensional space $\mathbb{R}^n$ by a random matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n\times N}$. The class of random matrices we discussed in this paper, defined below, are the ones with so-called JL Property.

Definition 1 [14] *A random matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n\times N}$ is said to satisfy JL Property if for any $0 < \varepsilon < 1/2$, there exists some universal constant $c > 0$, such that for any $\boldsymbol{x} \in \mathbb{R}^N$,*

$$\left|\|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| < \varepsilon\|\boldsymbol{x}\|_2^2 \tag{1}$$

holds with probability at least $1 - \exp(-c\varepsilon^2 n)$.

Random matrices with JL Property include Gaussian matrices, Bernoulli matrices, partial Fourier matrices and partial Hadamard matrices, which are easy to generate and efficient to compute. These random matrices have been revealed to be able to preserve plenty of connections between subspaces with overwhelming probability, such as the distance between subspaces and the canonical angles [13], [15], [16]. These properties suggest that we make almost no perturbation on the data structure after applying random matrices with JL Property.

JL Property also implies other natures of random matrices. The next lemma shows that the spectral norm of $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}$ is around one with high probability, which will be used in the proof of our main result.

Lemma 1 ([17], Theorem 5.39) *Suppose $\boldsymbol{\Phi} \in \mathbb{R}^{n\times N}$ is a random matrix which satisfies JL Property. Then for any $0 < \varepsilon < 1/2$, there exist universal constants $c_1, c_2 > 0$, such that for any $n > c_1\varepsilon^{-2}$,*

$$\left|\|\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\|_{2\to 2} - 1\right| < \varepsilon \tag{2}$$

holds with probability at least $1 - \exp(-c_2\varepsilon^2 n)$, where $\|\cdot\|_{2\to 2}$ denotes the spectral norm.

B. Self-representation-based OD

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_M] \in \mathbb{R}^{N\times M}$ denote the matrix whose columns are M data points with unit euclidian norm. Some of the data points are inliers lying on a union of m subspaces $\mathcal{S}_1, \dots, \mathcal{S}_m$, while the others are outliers which do not lie around any subspace.

Self-representation-based OD algorithm consists of two steps [10]. In the first step, a representation matrix, denoted as $\boldsymbol{R} = [\boldsymbol{r}_1, \dots, \boldsymbol{r}_M]$, is established by solving

$$\begin{aligned}\boldsymbol{r}_j &= \operatorname{argmin}_{\boldsymbol{c}} F(\boldsymbol{c}; \boldsymbol{x}_j, \boldsymbol{X}) \\ &\coloneqq \lambda\|\boldsymbol{c}\|_1 + \tfrac{1-\lambda}{2}\|\boldsymbol{c}\|_2^2 + \tfrac{\gamma}{2}\left\|\boldsymbol{x}_j - \boldsymbol{X}\boldsymbol{c}\right\|_2^2, \text{ s.t. } r_{jj} = 0,\end{aligned} \tag{3}$$

where $F(\cdot;\cdot,\cdot)$ is defined for simplicity and $\lambda \in [0,1)$, $\gamma > 0$ are parameters. From the representation matrix $\boldsymbol{R}$, a directed representation graph can be constructed. This graph has M nodes $V_1, \dots, V_M$, respectively representing M data points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$, and the weight of the edge from V_i to V_j is defined as $e_{ij} = |r_{ij}|$. The expected case is that all nodes corresponding to inliers have edges that only lead to inliers, while edges from outlier nodes can lead to both inliers and outliers. In this case, the corresponding representation matrix is said to satisfy the SP property, whose mathematical definition is below.

Definition 2 [10] Denote ℓ_j *as the index of the subspace data $\boldsymbol{x}_j$ lies in, i.e., $\boldsymbol{x}_j \in \mathcal{S}_{\ell_j}$, and $\boldsymbol{X}_j^{\ell_j}$ as the submatrix of $\boldsymbol{X}$ whose columns are all data in $\mathcal{S}_{\ell_j}$ except $\boldsymbol{x}_j$. Define $\boldsymbol{r}_j^{\ell_j} = \operatorname{argmin}_{\boldsymbol{c}} F\left(\boldsymbol{c}; \boldsymbol{x}_j, \boldsymbol{X}_j^{\ell_j}\right)$ and $\boldsymbol{\delta}_j \coloneqq \gamma\left(\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\boldsymbol{r}_j^{\ell_j}\right)$. If*

$$\max_j \max_{k\neq j, \boldsymbol{x}_k \in \mathcal{S}_{\ell_j}} \left|\langle \boldsymbol{x}_k, \boldsymbol{\delta}_j\rangle\right| < \lambda, \tag{4}$$

is satisfied, where λ is the parameter in $F(\cdot;\cdot,\cdot)$, then the representation matrix $\boldsymbol{R}$ of $\boldsymbol{X}$ is said to satisfy SP property.

In the second step, random walks are initialized on the representation graph from each node. When the SP property holds, all random walks eventually end up at inlier nodes, which is our criteria to separate inliers and outliers [10].

The correctness of self-representation-based OD is theoretically analyzed (shown in Lemma 1) under Assumption 1.

Assumption 1 [10] *In the representation graph, for any inlier subspace there is a path in each direction between each pair of*

vertices in the connected component, while for each subset of outliers there exists an edge that goes from one point in this subset to an inlier or to another outlier outside this subset.

Lemma 2 ([18], Lemma A.1) *If the representation matrix* $\boldsymbol{R}$ *satisfies Assumption 1, and the dataset* $\boldsymbol{X}$ *satisfy (4) (i.e.* $\boldsymbol{R}$ *satisfies SP property), then self-representation-based OD correctly identifies all outliers.*

III. Dimensionality-reduced Secure Outlier Detection

To tackle the problem of OD's poor security and high computational cost, we propose a new algorithm called DrSOD in Algorithm 1.

Algorithm 1 Dimensionality-reduced Secure Outlier Detection (DrSOD)

Input: Dataset $\boldsymbol{X} \in \mathbb{R}^{N\times M}$, compressed dimension n, number of iterations T, threshold ζ, parameters λ and γ in the elastic net problem.

Step I: Dimensionality reduction.

1. Generate a random matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n\times N}$.
2. Compute dimensionality-reduced dataset $\widetilde{\boldsymbol{X}} = \boldsymbol{\Phi X}$.

Step II: Self-representation.

3. Compute representation matrix $\widetilde{\boldsymbol{R}}$ of $\widetilde{\boldsymbol{X}}$ using (5).

Step III: Random walks.

4. Compute transition matrix $\boldsymbol{P}$ from $\widetilde{\boldsymbol{R}}$ using $p_{ij} = \frac{|r_{ij}|}{\|\boldsymbol{r}_i\|_1}, \forall i,j \in \{1,\dots,M\}$.
5. Initialize $t = 0$, $\boldsymbol{\pi} = [1/M,\dots,1/M]$, and $\overline{\boldsymbol{\pi}} = \boldsymbol{0}$.
6. **for** $t = 1,\dots,T$ **do**
7. Compute $\boldsymbol{\pi} \leftarrow \boldsymbol{\pi}\cdot\boldsymbol{P}$.
8. Set $\overline{\boldsymbol{\pi}} \leftarrow \overline{\boldsymbol{\pi}} + \boldsymbol{\pi}$.
9. **end for**
10. Set $\overline{\boldsymbol{\pi}} = \overline{\boldsymbol{\pi}}/T$.

Output: Indicator of outliers $\mathbb{I}(\overline{\boldsymbol{\pi}}_j < \zeta)$.

The first step of DrSOD is a preprocessing step which compresses dataset $\boldsymbol{X}$ by a random matrix $\boldsymbol{\Phi}$ and yields the dimensionality-reduced dataset $\widetilde{\boldsymbol{X}} = \boldsymbol{\Phi X}$ with every data point $\boldsymbol{x}_j$ compressed to $\widetilde{\boldsymbol{x}}_j = \boldsymbol{\Phi x}_j$. In the second and the third step, we apply self-representation-based OD algorithm [10] to the dimensionality-reduced dataset $\widetilde{\boldsymbol{X}}$. Specifically, we construct representation matrix $\widetilde{\boldsymbol{R}} = [\tilde{\boldsymbol{r}}_1,\dots,\tilde{\boldsymbol{r}}_M]$ by solving

$$\tilde{\boldsymbol{r}}_j = \operatorname{argmin}_{\boldsymbol{c}} F\left(\boldsymbol{c}; \widetilde{\boldsymbol{x}}_j, \widetilde{\boldsymbol{X}}\right), \tag{5}$$

where $F(\cdot;\cdot,\cdot)$ is defined in (3). Then we apply random walks on the representation graph $\widetilde{\boldsymbol{R}}$ to identify outliers.

Our main result of analyzing the correctness of DrSOD is shown in the following Theorem.

Theorem 1 *For any* $0 < \varepsilon < 1/4$*, if matrix* $\widetilde{\boldsymbol{R}}$ *defined in (5) satisfies Assumption 1, and the dataset* $\boldsymbol{X}$ *satisfies*

$$\max_j \max_{k\neq j, \boldsymbol{x}_k \in \mathcal{S}_{\ell_j}} \left|\langle \boldsymbol{x}_k, \boldsymbol{\delta}_j\rangle\right| < \lambda - \tau, \tag{6}$$

where $\boldsymbol{\delta}_j$ *is defined in Definition 2 and*

$$\tau \coloneqq \max_j \left\{\varepsilon\left(1 + \|\boldsymbol{\delta}_j\|_2^2\right) + \sqrt{6\varepsilon}(1+\varepsilon)\gamma\|\boldsymbol{\delta}_j\|_2\right\},$$

then there exist universal constants $c_1, c_2 > 0$*, such that for any* $n > \max(c_1\varepsilon^{-2}, \ln M)$*, DrSOD correctly identifies all outliers with probability at least* $1 - \exp(-c_2\varepsilon^2 n)$.

The holding probability increases exponentially with the compressed dimension n. This indicates that if n is not too small, then the probability is close to one.

Now we are ready to compare the condition (6) in Theorem 1 with (4) in the result of the self-representation-based OD Algorithm. Notice that (6) is only strengthened with an amount $\tau = \mathcal{O}\left(\varepsilon \max_j \|\boldsymbol{\delta}_j\|_2\right)$, which is determined by ε and can be sufficiently small. Thus, there is almost no excessive requirement on the raw dataset. Another difference in the condition of Theorem 1 and Lemma 2 is that the former requires $\widetilde{\boldsymbol{R}}$, instead of $\boldsymbol{R}$, to satisfy Assumption 1. The reasonability of this condition is explained in Remark 1.

A. Proof of Theorem 1

Because of the independence of $\{\boldsymbol{x}_j\}_{j=1}^M$, we complete the proof by taking fixed j and k which satisfy $\boldsymbol{x}_k \in \mathcal{S}_{\ell_j}$. Despite of the notations in Theorem 1 and Lemma 2, we denote $f(\boldsymbol{c}) \coloneqq F\left(\boldsymbol{c}; \boldsymbol{x}_j, \boldsymbol{X}_j^{\ell_j}\right)$ and $\tilde{f}(\boldsymbol{c}) \coloneqq F\left(\boldsymbol{c}; \widetilde{\boldsymbol{x}}_j, \widetilde{\boldsymbol{X}}_j^{\ell_j}\right)$, where $\widetilde{\boldsymbol{X}}_j^{\ell_j} = \boldsymbol{\Phi}\boldsymbol{X}_j^{\ell_j}$. Because $\tilde{f}(\boldsymbol{c})$ is strongly convex, it has unique optimal solution $\tilde{\boldsymbol{r}}_j^{\ell_j}$ and define $\widetilde{\boldsymbol{\delta}}_j \coloneqq \gamma(\widetilde{\boldsymbol{x}}_j - \widetilde{\boldsymbol{X}}_j^{\ell_j}\tilde{\boldsymbol{r}}_j^{\ell_j})$.

What we need to prove now is

$$\left|\langle \boldsymbol{x}_k, \boldsymbol{\delta}_j\rangle - \langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{\delta}}_j\rangle\right| \le \varepsilon\left(1 + \|\boldsymbol{\delta}_j\|_2^2\right) + \sqrt{6\varepsilon}(1+\varepsilon)\gamma\|\boldsymbol{\delta}_j\|_2,$$

holds with probability at least $1 - \exp(-C\varepsilon^2 n)$, where C is a universal constant. If this inequality holds, using union bound and $n > \ln M$, we can obtain that DrSOD satisfies SP property with probability at least $1 - M\exp(-C\varepsilon^2 n) > 1 - \exp(-c\varepsilon^2 n)$ and complete the proof.

We can first estimate $\left|\langle \boldsymbol{x}_k, \boldsymbol{\delta}_j\rangle - \langle \boldsymbol{\Phi x}_k, \boldsymbol{\Phi\delta}_j\rangle\right|$ by using a consequence drawn directly from the definition of JL Property.

Lemma 3 *Suppose* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$*, and* $\boldsymbol{\Phi} \in \mathbb{R}^{n\times N}$ *is a random matrix which satisfies JL Property. Denote* $\widetilde{\boldsymbol{x}} \coloneqq \boldsymbol{\Phi x}$ *and* $\widetilde{\boldsymbol{y}} \coloneqq \boldsymbol{\Phi y}$ *as the projected vectors. Then for any* $0 < \varepsilon < 1/2$*, there exists some universal constant* $c' > 0$*, such that*

$$|\langle \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle| < \varepsilon(\|\boldsymbol{x}\|_2^2 + \|\boldsymbol{y}\|_2^2), \tag{7}$$

holds with probability at least $1 - \exp(-c'\varepsilon^2 n)$.

PROOF OF LEMMA 3 Note that $\|\boldsymbol{a}+\boldsymbol{b}\|_2^2 - \|\boldsymbol{a}-\boldsymbol{b}\|_2^2 = 4\langle \boldsymbol{a}, \boldsymbol{b}\rangle$, we obtain

$$\begin{aligned}
&|\langle \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle| \\
&\le \frac{1}{4}\left|\|\widetilde{\boldsymbol{x}} + \widetilde{\boldsymbol{y}}\|_2^2 - \|\boldsymbol{x}+\boldsymbol{y}\|_2^2\right| + \frac{1}{4}\left|\|\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{y}}\|_2^2 - \|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right| \\
&< \frac{1}{4}\varepsilon\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 + \frac{1}{4}\varepsilon\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\
&= \frac{1}{2}\varepsilon(\|\boldsymbol{x}\|_2^2 + \|\boldsymbol{y}\|_2^2)
\end{aligned}$$

holds with probability at least $1 - 2\exp(-c\varepsilon^2 n)$. ∎

Using Lemma 3, we obtain

$$\left|\langle \boldsymbol{x}_k, \boldsymbol{\delta}_j\rangle - \langle \widetilde{\boldsymbol{x}}_k, \boldsymbol{\Phi}\boldsymbol{\delta}_j\rangle\right| < \varepsilon\left(1 + \left\|\boldsymbol{\delta}_j\right\|_2^2\right)$$

holds with probability at least $1 - 2\exp(-c'\varepsilon^2 n)$, where c' is a universal constant. Now we just need to prove

$$\left|\langle \widetilde{\boldsymbol{x}}_k, \boldsymbol{\Phi}\boldsymbol{\delta}_j\rangle - \langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{\delta}}_j\rangle\right| < \sqrt{6\varepsilon}(1+\varepsilon)\gamma\left\|\boldsymbol{\delta}_j\right\|_2$$

holds with probability at least $1 - 2\exp(-c''\varepsilon^2 n)$.

We compute the difference between $f\left(\boldsymbol{r}_j^{\ell_j}\right)$ and $f\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right)$ using Definition 1. Note that $\boldsymbol{c} = \boldsymbol{r}_j^{\ell_j}$ and $\boldsymbol{c} = \tilde{\boldsymbol{r}}_j^{\ell_j}$ are the optimal points of $f(\boldsymbol{c})$ and $\tilde{f}(\boldsymbol{c})$, respectively, we obtain $f\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right) \geq f\left(\boldsymbol{r}_j^{\ell_j}\right)$ and $\tilde{f}\left(\boldsymbol{r}_j^{\ell_j}\right) \geq \tilde{f}\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right)$. Then, we have

$$\begin{aligned} & f\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right) - f\left(\boldsymbol{r}_j^{\ell_j}\right) \\ & \leq \left|f\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right) - \tilde{f}\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right)\right| + \left|\tilde{f}\left(\boldsymbol{r}_j^{\ell_j}\right) - f\left(\boldsymbol{r}_j^{\ell_j}\right)\right| \\ & < \frac{\varepsilon\gamma}{2}\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\tilde{\boldsymbol{r}}_j^{\ell_j}\right\|_2^2 + \frac{\varepsilon\gamma}{2}\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\boldsymbol{r}_j^{\ell_j}\right\|_2^2 \end{aligned} \tag{8}$$

holds with probability at least $1 - 2\exp(-c\varepsilon^2 n)$.

Denote $\nabla\|\boldsymbol{c}\|_1 = \operatorname{sign}(\boldsymbol{c})$, where the function sign is defined componentwise by $\left(\operatorname{sign}(\boldsymbol{c})\right)_i = \operatorname{sign}(c_i)$ for nonzero c_i while $\operatorname{sign}(0) \in [-1,1]$, following the definition in [19].

We have

$$\begin{aligned} \left\|\tilde{\boldsymbol{r}}_j^{\ell_j}\right\|_1 - \left\|\boldsymbol{r}_j^{\ell_j}\right\|_1 &= \nabla\left\|\boldsymbol{r}_j^{\ell_j}\right\|_1^{\top}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right) \\ &+ \sum_{i=1}^{\dim \mathcal{S}_{\ell_j}-1}\left[\operatorname{sign}\left(\tilde{r}_{ji}^{\ell_j}\right)\tilde{r}_{ji}^{\ell_j} - \operatorname{sign}\left(r_{ji}^{\ell_j}\right)\tilde{r}_{ji}^{\ell_j}\right], \end{aligned}$$

and using $\operatorname{sign}\left(\tilde{r}_{ji}\right)\tilde{r}_{ji} \geq \operatorname{sign}(r_{ji})\tilde{r}_{ji}$, we obtain

$$\left\|\tilde{\boldsymbol{r}}_j^{\ell_j}\right\|_1 - \left\|\boldsymbol{r}_j^{\ell_j}\right\|_1 \geq \nabla\left\|\boldsymbol{r}_j^{\ell_j}\right\|_1^{\top}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right). \tag{9}$$

Using the convexity of $f(\boldsymbol{c})$ and $\nabla f\left(\boldsymbol{r}_j^{\ell_j}\right) = \boldsymbol{0}$, we have

$$\begin{aligned} & f\left(\tilde{\boldsymbol{r}}_j^{\ell_j}\right) - f\left(\boldsymbol{r}_j^{\ell_j}\right) \\ & \geq \nabla f\left(\boldsymbol{r}_j^{\ell_j}\right)^{\top}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right) \\ & \quad + \frac{1}{2}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)^{\top}\nabla^2 f\left(\boldsymbol{r}_j^{\ell_j}\right)\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right) \\ & = \frac{1-\lambda}{2}\left\|\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right\|_2^2 + \frac{1}{2}\gamma\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2 \\ & \geq \frac{1}{2}\gamma\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2. \end{aligned}$$

Moreover, using $\|\boldsymbol{a} - \boldsymbol{b}\|_2^2 \leq 2(\|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2)$ to handle $\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\tilde{\boldsymbol{r}}_j^{\ell_j}\right\|_2^2$ in (8), we obtain

$$\begin{aligned} & \left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2 \\ & < \varepsilon\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\boldsymbol{r}_j^{\ell_j}\right\|_2^2 + \varepsilon\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\tilde{\boldsymbol{r}}_j^{\ell_j}\right\|_2^2 \\ & \leq 3\varepsilon\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\boldsymbol{r}_j^{\ell_j}\right\|_2^2 + 2\varepsilon\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2, \end{aligned}$$

or equivalently we have

$$\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2 < \frac{3\varepsilon}{1-2\varepsilon}\left\|\boldsymbol{x}_j - \boldsymbol{X}_j^{\ell_j}\boldsymbol{r}_j^{\ell_j}\right\|_2^2$$

holds with probability at least $1 - 2\exp(-c\varepsilon^2 n)$.

Furthermore, using Lemma 1, we have

$$\begin{aligned} \left|\langle \widetilde{\boldsymbol{x}}_k, \boldsymbol{\Phi}\boldsymbol{\delta}_j\rangle - \langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{\delta}}_j\rangle\right| &= \gamma\left|\left\langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{X}}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\rangle\right| \\ &= \gamma\left|\left\langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{X}}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\rangle\right| \\ &\leq \gamma\|\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi}\|_{2\to 2}\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2 \\ &\leq \gamma(1+\varepsilon)\sqrt{\frac{3\varepsilon}{1-2\varepsilon}}\left\|\boldsymbol{\delta}_j\right\|_2 \end{aligned}$$

holds with probability at least $1 - \exp(-c''\varepsilon^2 n)$, where c'' is a universal constant.

Note that $1 - 2\varepsilon > 1/2$, we complete the proof. ∎

Remark 1 *The assumption of the representation graph needed in Theorem 1 is the same as in Lemma 2. This is because* $\boldsymbol{r}_j$ *and* $\tilde{\boldsymbol{r}}_j$ *are very similar. Specifically, we can obtain that*

$$\left\|\tilde{\boldsymbol{r}}_j - \boldsymbol{r}_j\right\|_2^2 < \frac{\varepsilon\gamma}{1-\lambda}\left(\left\|\boldsymbol{x}_j - \boldsymbol{X}\boldsymbol{r}_j\right\|_2^2 + \left\|\boldsymbol{x}_j - \boldsymbol{X}\tilde{\boldsymbol{r}}_j\right\|_2^2\right)$$

also holds with probability at least $1 - 2\exp(-c\varepsilon^2 n)$*, and the derivation is similar to* $\left\|\boldsymbol{X}_j^{\ell_j}\left(\tilde{\boldsymbol{r}}_j^{\ell_j} - \boldsymbol{r}_j^{\ell_j}\right)\right\|_2^2$.

B. Complexity Analysis

The step accounts for the large computation complexity is the optimization in (5), and there are many ways to solve it. We take a typical method RSSN [19] as an example. For an $N \times M$-dimensional dataset, it takes $\mathcal{O}(N^2M^3)$ time for one iteration, so when the dimension of the dataset is reduced to $n \times M$, the running time in self-representation reduces significantly, while the additional cost is minor as the time complexity of dimensionality reduction is just $\mathcal{O}(N \log N\, M^2)$.

IV. Experimental Results

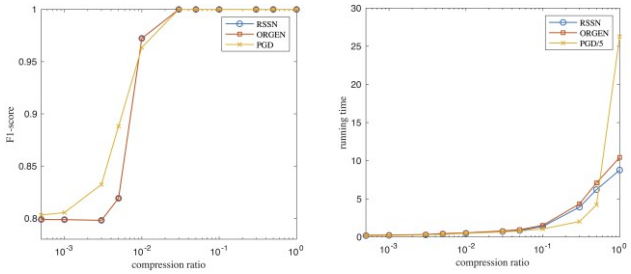
In this section, we further verify the effectiveness and efficiency of DrSOD by conducting experiments on a synthetic dataset and a real-world dataset. The efficiency is shown directly by the significant reduction in running time. The metric we provide for describing the practicability is the F1-score [10], which is the harmonic mean of precision and recall. We report the largest F1-score across all thresholds, so when an F1-score is 1, there exists a threshold that separates inliers and outliers perfectly. On the real-world dataset, we also use visible images to show the security and privacy consideration of DrSOD. The parameters in (5) are set to $\lambda = 0.95$ and $\gamma = \alpha\lambda/\max_{i:i\neq j}|\boldsymbol{x}_j^{\top}\boldsymbol{x}_i|$, where α is respectively 100 and 10 in the following experiments. The number of iterations in Algorithm 1 is $T = 10$.

We choose three methods to solve the convex optimization problem (5): Regularized SemiSmooth Newton method (RSSN) [19], ORacle Guided Elastic Net method (ORGEN) [18], and Proximal Gradient Decent method (PGD) [20]. The practicability and efficiency vary according to the parameters and the complexity of the chosen method. For instance, PGD is proved to reach the optimal solution but has a larger

computational complexity compared to RSSN and ORGEN. However, they have all shown the same tendency which coincide with the conclusion in our theoretical analysis, i.e., the F1-score preserves when the compression ratio is not too small, and the running time declines significantly as the compression ratio becomes lower.

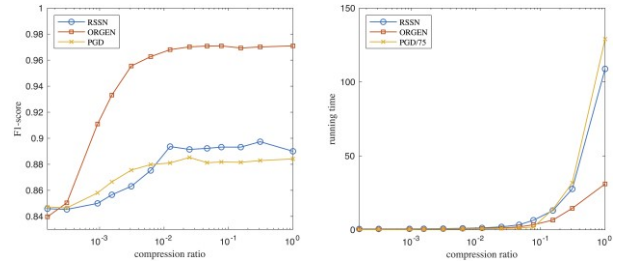
A. DrSOD on Synthetic Dataset

The synthetic dataset is constructed in the ambient space of dimension 500. We randomly choose two subspaces of dimension 10 and generate 50 points in two subspaces as inliers respectively, and the outliers are 50 arbitrary points in the ambient space. Fig. 2(a) and (b) show the F1-score and the running time under different compression ratios respectively. It can be seen that when using ORGEN to solve the optimization problem (5), the detection still succeeds when the compression ratio decreases from 1 to 0.03, while the running time drops from 10.43 to 0.80. Apparently, the correctness completely preserves and considerable time is saved after compression in this model.



(a) F1-score (b) running time

Fig. 2. Performance of DrSOD on synthetic dataset. (a) and (b) show that OD still succeeds when the compression ratio is 0.03, while the running time drops to about 8% of the original algorithm.



(a) F1-score (b) running time

Fig. 3. Performance of DrSOD on real-world dataset. (a) and (b) show that the F1-score is almost unchanged when the compression ratio is greater than 0.025, while the running time declines significantly.

B. DrSOD on Real-world Dataset

For real-world dataset, we use standard dataset Extended Yale Face B [21], consisting of images of 38 individuals under 64 different lighting conditions, and the resolution of each image is 192×168. We conduct an experiment aiming at identify the images of people who appear less among a batch of images. We select all 64 images of the first 3 individuals in Extended Yale Face B as inliers and 2 images of the other individuals chosen randomly as outliers, i.e., 262 data points are used, with 70 outliers expected to be detected. In Fig. 3(a) and (b), take ORGEN as an example, on one hand, the F1-score using the original 32256-dimensional images is 0.9710, and even when the images are compressed to 800-dimensional, the F1-score is still 0.9702, which is an ignorable decrease. On the other hand, the saving in running time is significant, which goes from 31.02 to 1.23, so the compression ratio over 0.025 is demonstrated to achieve satisfactory results.

We use Fig. 1 to illustrate the security and privacy consideration of DrSOD. A random sample of the raw data is shown in Fig. 1(a). Previous methods need direct access to these visible images, which is harmful to data security and privacy. However, in DrSOD, people who process the same image shown in Fig. 1(a) just need the preprocessed information shown in Fig. 1(b) (take compression ratio 0.0625 as an example), which is safer than previous methods.

V. CONCLUSION

In this paper, we propose DrSOD algorithm which could protect the data privacy and has high computational efficiency. Moreover, we theoretically confirm the correctness of DrSOD. The experiments on synthetic and real-world datasets justify the superior performance of detection and the significant reduction on computational cost.

REFERENCES

[1] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 10, pp. 1832–1845, 2009.

[2] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pp. 218–233, 2003.

[3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2765–2781, 2013.

[4] B. Mcwilliams and G. Montana, "Subspace clustering of highdimensional data: a predictive approach," Data Mining and Knowledge Discovery, vol. 28, no. 3, pp. 736–772, 2014.

[5] D. Zhang, "Big data security and privacy protection," International Conference on Management and Computer Science, 2018.

[6] G. Lerman and T. Zhang, "Robust recovery of multiple subspaces by geometric opminimization," Annals of Statistics, vol. 39, no. 5, pp. 2686–2715, 2011.

[7] R. Heckel and H. Bolcskei, "Robust subspace clustering via thresholding," IEEE Transactions on Information Theory, vol. 61, no. 11, pp. 6320–6342, 2015.

[8] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," Advances in Neural Information Processing systems, pp. 2080–2088, 2009.

[9] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," Advances in Neural Information Processing systems, pp. 2496–2504, 2010.

[10] C. You, D. P. Robinson, and R. Vidal, "Provable self-representation based outlier detection in a union of subspaces," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4323–4332, 2017.

[11] G. Li, Q. Liu, and Y. Gu, "Rigorous restricted isometry property of low-dimensional subspaces," Applied and Computational Harmonic Analysis, pp. 608–635, 2020.

[12] L. Meng, G. Li, J. Yan, and Y. Gu, "A general framework for understanding compressed subspace clustering algorithms," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 6, pp. 1504–1519, 2018.

[13] X. Xu, G. Li, and Y. Gu, "Unraveling the veil of subspace rip through near-isometry on subspaces," accepted by IEEE Transactions on Signal Processing, available at arXiv:1905.09608, 2019.

[14] S. Foucart and H. Rauhut, “A mathematical introduction to compressive sensing,” Springer Science and Business Media, 2013.

[15] Y. Jiao, X. Shen, G. Li, and Y. Gu, “Subspace principal angle preserving property of gaussian random projection,” IEEE Data Science Workshop, pp. 115–119, 2018.

[16] G. Li and Y. Gu, “Restricted isometry property of gaussian random projection for finite set of subspaces,” IEEE Transactions on Signal Processing, vol. 66, no. 7, pp. 1705–1720, 2018.

[17] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” arXiv preprint arXiv:1011.3027, 2010.

[18] C. You, C. Li, D. P. Robinson, and R. Vidal, “Oracle based active set algorithm for scalable elastic net subspace clustering,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 3928–3937, 2016.

[19] B. Jin, D. A. Lorenz, and S. Schiffler, “Elastic-net regularization: error estimates and active set methods,” Inverse Problems, vol. 25, no. 11, p. 115022, 2009.

[20] N. Parikh, S. Boyd et al., “Proximal algorithms,” Foundations and Trends in Optimization, vol. 1, no. 3, pp. 127–239, 2014.

[21] K. Lee, J. Ho, and D. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684–698, 2005.