

# Restricted Isometry Property of Gaussian Random Projection for Finite Set of Subspaces

Gen Li and Yuantao Gu\*

April 10, 2017

## Abstract

Dimension reduction plays an essential role when decreasing the complexity of solving large-scale problems. The well-known Johnson-Lindenstrauss (JL) Lemma and Restricted Isometry Property (RIP) admit the use of random projection to reduce the dimension while keeping the Euclidean distance, which leads to the boom of Compressed Sensing and the field of sparsity related signal processing. Recently, successful applications of sparse models in computer vision and machine learning have increasingly hinted that the underlying structure of high dimensional data looks more like a union of subspaces (UoS). In this paper, motivated by JL Lemma and an emerging field of Compressed Subspace Clustering, we study for the first time the RIP of Gaussian random matrix for compressing two subspaces. We theoretically prove that with high probability the *affinity* or *distance* between two projected subspaces are concentrated around their estimates. When the ambient dimension after projection is sufficiently large, the affinity and distance between two subspaces almost remain unchanged after random projection. Numerical experiments verify the theoretical work.

**Keywords:** Johnson-Lindenstrauss Lemma, Restricted Isometry Property, Gaussian random matrix, union of subspaces, affinity, projection, compression

## 1 Introduction

In the big data era we confront with large-scale problems dealing with data points or features in high dimensional vector spaces. In the enduring effort of trying to decrease the complexity of solving such large problems, dimension reduction has played an essential role. The well-known Johnson-Lindenstrauss (JL) Lemma [1, 2] and the Restricted Isometry Property (RIP) [3, 4, 5] allow the use of random projection to reduce the space dimension

---

\*The authors are with the Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, CHINA. The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

A short version of this work has been presented at ICASSP 2017.

while keeping the Euclidean distance between any two data points, which leads to the boom of Compressed Sensing (CS) and the field of sparsity related signal processing [6, 7, 8, 9, 10].

Typically, the problem of CS is described as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x},$$

where  $\mathbf{x} \in \mathbb{R}^N$  is a  $k$ -sparse signal we wish to recover,  $\mathbf{y} \in \mathbb{R}^n, n < N$  is the available measurement, and  $\mathbf{\Phi} \in \mathbb{R}^{n \times N}$  is a known projection matrix. To sufficiently ensure robust recovery to the original signal, the projection matrix should approximately preserve the distance between any two  $k$ -sparse signals. Specifically, JL Lemma states that, for any set  $\mathcal{V}$  of  $L$  points in  $\mathbb{R}^N$ , if  $n$  is a positive integer such that

$$n \geq 4 \left( \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} \right)^{-1} \ln N,$$

then there exists a map  $f : \mathbb{R}^N \rightarrow \mathbb{R}^n$ , such that for all  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{V}$ ,

$$(1 - \varepsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|_2^2 \leq (1 + \varepsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2,$$

where  $0 < \varepsilon < 1$  is a constant. Moreover, RIP is a generalization of this lemma. We say that the projection matrix  $\mathbf{\Phi}$  satisfies RIP of order  $k$  with  $\delta_k$  as the smallest nonnegative constant, such that

$$(1 - \delta_k) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}_1 - \mathbf{\Phi}\mathbf{x}_2\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2.$$

holds for any two  $k$ -sparse vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

In CS, we generally construct the measurement matrix by selecting  $\mathbf{\Phi}$  as a random matrix. For example, we draw the matrix elements  $\phi_{ji}$  independently from Gaussian distribution  $\mathcal{N}(0, 1/n)$  [6, 7, 11]. More rigorously, using concentration of measure arguments [12, 13, 14],  $\mathbf{\Phi}$  is shown to have the RIP with high probability if  $n \geq ck \ln(N/k)$ , with  $c$  a small constant. In addition, there are theoretical results showing some angle-preserving properties as well [15, 16].

Furthermore, in [17, 18], the signals of interest have been extended from conventional sparse vectors to the vectors that belong to a union of subspaces (UoS). Nowadays, UoS becomes an important topic [18], and plays an important role in many subfields of CS, such as multiple measurement vector [19, 20] and block sparse recovery [21, 22]. It has been proved in [23, 24] that, with high probability the random projection matrix  $\mathbf{\Phi}$  can preserve the distance between two signals belonging to a UoS. Recently, the stable embedding property has been extended to signals modeled as low-dimensional Riemannian submanifolds in Euclidean space [25, 26, 27].

## 1.1 Motivation

Although the UoS model is very popular and is extensively used in various applications, few theoretical analysis describes the property of the random projection for linear subspaces.

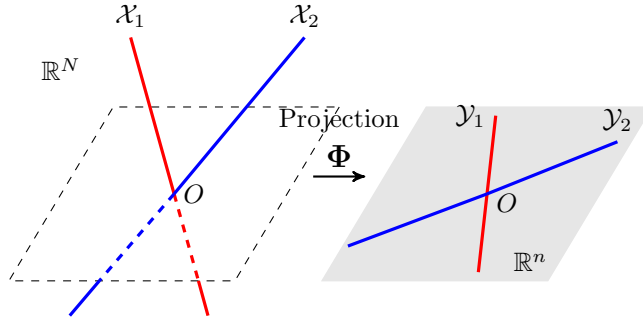


Figure 1: This paper studies the Restricted Isometry Property of Gaussian random projection matrices for a set of subspaces.

Recently, many researches have emerged, focusing on different aspects of Subspace Clustering (SC) [28, 29, 30, 31]. A ready approach to reduce the complexity of SC is to compress the original samples into low dimensional vectors and then to cluster by their subspaces in the new low dimensional space, which is called Compressed SC (CSC) or dimensionality-reduced SC [32, 33]. Based on the concept of affinity, which characterizes the similarity between two subspaces, it has been theoretically proved and numerically verified that several dominant algorithms could successfully perform clustering on the compressed data [34, 35].

This motivates our work to discover the distance-preserving property of random matrix for a set of subspaces, like what JL Lemma guarantees for finite signal set and RIP for sparse signals.

## 1.2 Main contribution

In this paper, motivated by the significance of JL Lemma and the feasibility of CSC, we study the RIP of Gaussian random matrices for projecting a set of finite subspaces. After introducing the projection  $F$ -norm distance and build a metric space of the set of low-dimensional subspaces, we reveal the connection between affinity and distance, and lays a solid foundation for this work. Then we start from a simple case that one subspace is of dimension one and prove the concentration of affinity after random projection. Consequently, the general case is studied that the subspaces are of arbitrary dimensions. Based on our recent work that column-wise normalization well approximates the Gram-Schmidt orthogonalization in high-dimensional scenario [36], we successfully reach the RIP of two subspaces. Finally, the main contribution is generalized to a finite set of subspaces, as stated in Theorem 1.

**Theorem 1** *For any set composed by  $L$  subspaces  $\mathcal{X}_1, \dots, \mathcal{X}_L \in \mathbb{R}^N$  of dimension no more than  $d$ , if they are projected into  $\mathbb{R}^n$  by a Gaussian random matrix  $\Phi \in \mathbb{R}^{n \times N}$ ,*

$$\mathcal{X}_k \xrightarrow{\Phi} \mathcal{Y}_k = \{\mathbf{y} | \mathbf{y} = \Phi \mathbf{x}, \forall \mathbf{x} \in \mathcal{X}_k\}, \quad k = 1, \dots, L,$$

and  $d \ll n < N$ , then we have

$$(1 - \varepsilon)D^2(\mathcal{X}_i, \mathcal{X}_j) \leq D^2(\mathcal{Y}_i, \mathcal{Y}_j) \leq (1 + \varepsilon)D^2(\mathcal{X}_i, \mathcal{X}_j), \quad \forall i, j$$

with probability at least

$$1 - \frac{2dL(L-1)}{(\varepsilon - d/n)^2n},$$

when  $n$  is large enough, where  $D(\cdot)$  denotes the projection  $F$ -norm distance between two subspaces.

Although different metrics and distance measures have been used to describe the topological structure of the Grassmann manifold [37, 38, 39], as far as we know, there is no rigorous theoretical analysis for the RIP regarding to subspaces. This paper theoretically studies this problem for the first time. Numerical simulations are also provided to validate the theoretical results.

## 2 Problem Formulation

In this paper, we study the separability of subspaces after random projection. We first introduce the principal angles to describe the relative position and the affinity to measure the similarity between two subspaces. Considering that the relation of subspaces reflected by affinity does not possess good features of metric space, we then introduce the projection  $F$ -norm distance for evaluating the separability of subspaces. To be highlighted, we discover the connection between affinity and the above distance, which lays the foundation for the main contribution of this work.

### 2.1 Principal angles and affinity

The principal angles (or canonical angles) between two subspaces provide the best way to characterize the relative subspace positions. It has been introduced by Jordan [40] in 1875 and then rediscovered for several times. One may read [41] and the references herein for more usages of principal angles.

**Definition 1** *The principal angles  $\theta_1, \dots, \theta_{d_1}$  between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of dimensions  $d_1 \leq d_2$ , are recursively defined as*

$$\cos \theta_i = \max_{\mathbf{x}_1 \in \mathcal{X}_1} \max_{\mathbf{x}_2 \in \mathcal{X}_2} \frac{\mathbf{x}_1^\top \mathbf{x}_2}{\|\mathbf{x}_1\| \|\mathbf{x}_2\|} =: \frac{\mathbf{x}_{1i}^\top \mathbf{x}_{2i}}{\|\mathbf{x}_{1i}\| \|\mathbf{x}_{2i}\|}, \quad (1)$$

with the orthogonality constraints  $\mathbf{x}_k^\top \mathbf{x}_{kj} = 0, j = 1, \dots, i-1, k = 1, 2$ .

An alternative way of computing principal angles is to use the singular value decomposition [42].

**Lemma 1** *Let the columns of  $\mathbf{U}_k$  be orthonormal bases for subspace  $\mathcal{X}_k$  of dimension  $d_k, k = 1, 2$  and suppose  $d_1 \leq d_2$ . Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{d_1} \geq 0$  be the singular values of  $\mathbf{U}_1^T \mathbf{U}_2$ , then*

$$\cos \theta_i = \lambda_i, \quad i = 1, \dots, d_1.$$

When studying the problem of subspace clustering, affinity has been defined by utilizing principal angles to measure the subspace similarity [29].

**Definition 2** *The affinity between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of dimension  $d_1 \leq d_2$  is defined as*

$$\text{aff}(\mathcal{X}_1, \mathcal{X}_2) := \left( \sum_{i=1}^{d_1} \cos^2 \theta_i \right)^{\frac{1}{2}}. \quad (2)$$

Using Lemma 1 in Definition 2, we may readily introduce an algebraic approach for calculating affinity as follows.

**Lemma 2** *The affinity between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  can be calculated by*

$$\text{aff}(\mathcal{X}_1, \mathcal{X}_2) := \|\mathbf{U}_1^T \mathbf{U}_2\|_F, \quad (3)$$

where the columns of  $\mathbf{U}_k$  are orthonormal bases of  $\mathcal{X}_k, k = 1, 2$ .

## 2.2 From similarity to distance

In order to take advantage of the properties of metric space, we prefer to study the relation between subspaces by distances. Actually, there are several different definitions of distance based on the principal angles between two subspaces [43]. In this work, we focus on the projection  $F$ -norm distance. When two subspaces are of the same dimension, this distance is defined as follows.

**Definition 3** *The projection  $F$ -norm distance between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of the same dimension  $d$  is defined as*

$$D_1(\mathcal{X}_1, \mathcal{X}_2) := \frac{1}{\sqrt{2}} \|\mathbf{P}_1 - \mathbf{P}_2\|_F = \left( \sum_{i=1}^d \sin^2 \theta_i \right)^{\frac{1}{2}},$$

where  $\mathbf{P}_k = \mathbf{U}_k \mathbf{U}_k^T$  denotes the projection matrix for subspace  $\mathcal{X}_k, k = 1, 2$  and  $\theta_i, 1 \leq i \leq d$  denote the principal angles between the two subspaces.

When the dimensions of the two subspaces are different, which is the case of wide applications, we may accordingly generate the projection  $F$ -norm distance as follows.

**Definition 4** *The projection  $F$ -norm distance between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of dimension  $d_1, d_2$  is defined as*

$$D(\mathcal{X}_1, \mathcal{X}_2) := \frac{1}{\sqrt{2}} \|\mathbf{P}_1 - \mathbf{P}_2\|_F, \quad (4)$$

where  $\mathbf{P}_k = \mathbf{U}_k \mathbf{U}_k^T$  denotes the projection matrix for subspace  $\mathcal{X}_k, k = 1, 2$ .

One may readily check that this definition meets all requirements in the definition of distance measure, i.e. non-negativity, positive-definiteness, symmetry, and triangular inequality, thus the space of different dimensional subspaces becomes a metric space. This could also be derived by deeming  $\mathcal{X}_k$  as points on Grassmann manifold [44].

**Remark 1** *We want to stress that (4) is different from  $\left(\sum_{i=1}^{d_1} \sin^2 \theta_i\right)^{1/2}$ , which does not possess positive-definiteness, therefore violates the definition of a distance. This implies that principal angles are not sufficient for inducing a distance metric.*

Combining Definition 2 and Definition 4, we reveal for the first time the relationship between distance and affinity.

**Lemma 3** *The distance and affinity between two subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of dimension  $d_1, d_2$ , are connected by*

$$D^2(\mathcal{X}_1, \mathcal{X}_2) = \frac{d_1 + d_2}{2} - \text{aff}^2(\mathcal{X}_1, \mathcal{X}_2). \quad (5)$$

PROOF The proof is postponed to Appendix 7.1. ■

Because of the concise definition and easy computation of affinity, in this work we always start the theoretical analysis with affinity, and then present the results with distance by using Lemma 3. In addition, many discussions are also conducted based on the concept of affinity.

### 2.3 Projection of subspaces

We will focus on the RIP of randomly projecting two low-dimensional subspaces from a high-dimensional ambient space to a medium-dimensional ambient space.

**Definition 5** *Let  $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^N$  be two subspaces of dimension  $d_1 \leq d_2 \ll N$ . They are randomly projected to  $\mathbb{R}^n, d_2 \ll n < N$  as  $\mathcal{Y}_k$ ,*

$$\mathcal{X}_k \xrightarrow{\Phi} \mathcal{Y}_k = \{\mathbf{y} | \mathbf{y} = \Phi \mathbf{x}, \forall \mathbf{x} \in \mathcal{X}_k\}, \quad k = 1, 2.$$

where the projection matrix  $\Phi \in \mathbb{R}^{n \times N}$ ,  $n < N$ , is composed of entries independently drawn from Gaussian distribution  $\mathcal{N}(0, 1/n)$ .

**Remark 2** *Because  $d_1 \leq d_2 < n$ , one may notice that the dimension of subspaces remains unchanged after random projection in statistical sense.*

Following the above definition, we will study the change of the distance caused by random projection. For simplifying notation, we denote  $D_{\mathcal{X}} = D(\mathcal{X}_1, \mathcal{X}_2)$  and  $D_{\mathcal{Y}} = D(\mathcal{Y}_1, \mathcal{Y}_2)$  as the distances before and after random projection. Similarly, we use  $\text{aff}_{\mathcal{X}} = \text{aff}(\mathcal{X}_1, \mathcal{X}_2)$  and  $\text{aff}_{\mathcal{Y}} = \text{aff}(\mathcal{Y}_1, \mathcal{Y}_2)$  to denote the affinities before and after projection. Without loss of generality, we always suppose that  $d_1 \leq d_2$ . We call the affinity (distance) after random projection as *projected affinity (distance)*.

### 3 Main Results

In this section, we present our results about the RIP of subspaces after random projection. It begins with a simple case of estimating the projected affinity of a line and a subspace, and then the result is extended to the case of two subspaces with arbitrary dimensions. Finally, the RIP for subspaces is stated.<sup>1</sup>

#### 3.1 Concentration of the affinity between a line and a subspace after random projection

We first focus on a special case that one subspace is restricted to be a line (one-dimensional subspace). We begin from the concentration of their affinity caused by random projection and then replace the metric by the introduced distance.

The affinity between a line and a subspace will increase and concentrate on an estimate after Gaussian random projection. When the dimension of the new ambient space is large enough, the affinity almost remains unchanged after projection, as revealed in Lemma 4.

**Lemma 4** *Suppose  $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^N$  are a line and a  $d$ -dimension subspace,  $d \geq 1$ , respectively. Let  $\lambda = \text{aff}_{\mathcal{X}}$  denote their affinity. If they are projected onto  $\mathbb{R}^n, n < N$ , by a Gaussian random matrix  $\Phi \in \mathbb{R}^{n \times N}$ ,  $\mathcal{X}_k \xrightarrow{\Phi} \mathcal{Y}_k, k = 1, 2$ , then the affinity after projection,  $\text{aff}_{\mathcal{Y}}$ , can be estimated by*

$$\overline{\text{aff}}_{\mathcal{Y}}^2 = \lambda^2 + \frac{d}{n} (1 - \lambda^2), \quad (6)$$

where the estimation error is bounded by

$$\mathbb{P} \left( \left| \text{aff}_{\mathcal{Y}}^2 - \overline{\text{aff}}_{\mathcal{Y}}^2 \right| > \lambda^2 (1 - \lambda^2) \varepsilon \right) \leq \frac{4}{\varepsilon^2 n}, \quad (7)$$

when  $n$  is large enough.

PROOF The proof is postponed to Section 4.1. ■

As revealed in Lemma 4, the affinity between a line and a subspace increases after they are projected with the projection matrix specified as a Gaussian random matrix. Furthermore, the projected affinity can be estimated by (6) with high probability.

**Remark 3** *1) It's evident that the projected affinity will increase. The reason is that the angle between the line and the subspace decreases largely after they are projected from a high-dimensional space to a medium-dimensional space. 2) The increment caused by projection is in direct ratio to the dimension of the subspace. This means that the principal angle between the line and the subspace after projection drops more, when the subspace is of larger dimensionality. 3) The increment caused by projection is in inverse ratio to the dimension*

---

<sup>1</sup>Notice that the notation of *less than* in this work holds in the sense of equivalence. For example, if  $f(n) \leq 1/(n-2)$ , we may state that, without confusion,  $f(n) \leq 1/n$  when  $n$  is large enough for simplicity, considering that  $1/(n-2) \sim 1/n$ . We stress that the rigorousness of this work does not suffer.

of the ambient space after projection,  $n$ . When  $n$  is large enough, the affinity remains almost unchanged after projection. A visualization is that the smaller  $n$ , the larger probability that the line is contained within the subspace after projection, or the smaller the angle between them. 4) The increment of affinity is also determined by the affinity itself, which is  $[0, 1]$ . When the affinity is close to zero, which means the line is approximately orthogonal to the subspace, the increment caused by projection is the largest. On the contrary, when the affinity approaches one, which means the line is almost contained within the subspace, the increment is the smallest.

Now we will discuss the concentration of the projected affinity on its estimate, as shown in (7).

**Remark 4** 1) The probability that the projected affinity deviates from its estimate is below a threshold, which reduces by the rate of  $1/\varepsilon^2$ , where  $\varepsilon$  is in direct ratio to the accuracy of the estimate. This demonstrates that the projected affinity concentrate well on its estimate. 2) The threshold decreases to zero by the rate of  $1/n$  when the ambient dimension  $n$  increases. This means that in the high-dimensional scenario, the projected affinity concentrates on its estimate with a very large probability. Recalling Remark 3.3), one may conclude that the affinity remains unchanged with a large probability for high-dimensional problem. 3) With a given probability, the accuracy of estimation also depends on the original affinity. Consequently, the estimate is exactly accurate in two situations, where the line is almost contained within or orthogonal to the subspace.

Applying Lemma 3 in Lemma 4 to replace affinity by distance, we may readily reach the concentration of distance when randomly projecting a line and a subspace.

**Corollary 1** Let  $D_{\mathcal{X}}$  denote the distance between a line  $\mathcal{X}_1$  and a  $d$ -dimension subspace  $\mathcal{X}_2$ . The distance after projection,  $D_{\mathcal{Y}}$ , can be estimated by

$$\overline{D}_{\mathcal{Y}}^2 = D_{\mathcal{X}}^2 - \frac{d}{n} \left( D_{\mathcal{X}}^2 - \frac{d-1}{2} \right). \quad (8)$$

When  $n$  is large enough, the estimation error is bounded by

$$\mathbb{P} \left( \left| D_{\mathcal{Y}}^2 - \overline{D}_{\mathcal{Y}}^2 \right| > \lambda^2(1 - \lambda^2)\varepsilon \right) \leq \frac{4}{\varepsilon^2 n}. \quad (9)$$

When evaluating the impact of projection with *distance* instead of *affinity*, the changes of the estimate, the concentration, and their dependences on  $n, d$ , and the original metric are similar with those in Remark 3 and 4.

To sum up, we reveal that both affinity and distance between a line and a subspace concentrate on their estimates after random projection. By increasing the new ambient dimensionality, the metrics remain almost unchanged with a high probability.



### 3.2 Concentration of the affinity between two subspaces after random projection

we then study the general case of projecting two subspaces of arbitrary dimensions. Similar to the approach in the previous subsection, we begin with the concentration of affinity and then transform to distance.

The concentration of the affinity between two arbitrary subspaces after random projection are revealed in Theorem 2.

**Theorem 2** *Suppose  $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^N$  are two subspaces with dimension  $d_1 \leq d_2$ , respectively. Define*

$$\overline{\text{aff}}_{\mathcal{Y}}^2 = \text{aff}_{\mathcal{X}}^2 + \frac{d_2}{n}(d_1 - \text{aff}_{\mathcal{X}}^2) \quad (10)$$

*to estimate the affinity between two subspaces after random projection,  $\mathcal{X}_k \xrightarrow{\Phi} \mathcal{Y}_k, k = 1, 2$ . When  $n$  is large enough, the estimation error is bounded by*

$$\mathbb{P} \left( \left| \text{aff}_{\mathcal{Y}}^2 - \overline{\text{aff}}_{\mathcal{Y}}^2 \right| > \text{aff}_{\mathcal{X}}^2 \varepsilon \right) \leq \frac{4d_1}{\varepsilon^2 n}. \quad (11)$$

PROOF The proof is postponed to Section 4.2. ■

Recalling the discussions about projecting a line and a subspace in Remark 3, we may readily check that item 1) and 3) also hold in the situation of projecting two subspaces in Theorem 2. We will discuss the other two items in Remark 5.

**Remark 5** *1) The increment of affinity caused by projection is in direct ratio to the larger dimension of two subspaces. This comes from the fact that each basis of the lower dimensional subspace may be deemed as a one-dimensional subspace, the affinity between which and the higher dimensional subspace is evaluated. 2) The increment caused by projection is in direct ratio to  $(d_1 - \text{aff}_{\mathcal{X}}^2)$ . Notice that  $\text{aff}_{\mathcal{X}}^2 \in [0, d_1]$ . When  $\text{aff}_{\mathcal{X}}^2$  is close to zero, which means that two subspaces are almost orthogonal to each other, the increment is the largest and in direct ratio to  $d_1$ . When  $\text{aff}_{\mathcal{X}}^2$  is close to  $d_1$ , which means that the lower-dimensional subspace is almost contained within the other subspace, the increment must be the smallest among all situations.*

When  $d_1$  reduces to one, it's easy to check that the estimate of projected affinity (10) in Theorem 2 degenerates to (6) in Lemma 4. However, it is obvious that the bound of (11) can not reduce to (7). The reason is that the former is a loose result that undergoes much relaxation. Another version of Theorem 2 is given in Lemma 5, which is tight enough and exactly degenerates to Lemma 4.

**Lemma 5** *Following the same conditions and notations in Theorem 2, when  $n$  is large enough, the estimation error is bounded by*

$$\mathbb{P} \left( \left| \text{aff}_{\mathcal{Y}}^2 - \overline{\text{aff}}_{\mathcal{Y}}^2 \right| > \sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \varepsilon \right) \leq \frac{4d_1}{\varepsilon^2 n}, \quad (12)$$

where  $\lambda_i = \cos \theta_i$  and  $\theta_i$  denotes the principal angles between the original subspaces.

PROOF The proof is postponed to Section 4.2. ■

We keep both Theorem 2 and Lemma 5 deliberately as the main results, because they provide complementary usages. While (11) produces a clear formulation which is ready to calculate by using the original affinity as a whole, (12) reveals the relation between estimating accuracy and principal angles for us perceiving the intension.

Recalling Remark 4 about projecting a line and a subspace, the first two items both hold in the scenario of projecting two subspaces. We will recheck the last item.

**Remark 6** *With a given probability, the accuracy of estimation depends on all principal angles, i.e., in direct ratio to the sum of all  $\lambda_i^2(1 - \lambda_i^2)$ . This means that when two subspaces are almost orthogonal to each other or one is contained within the other, the estimate is accurate. The reason is as that in Remark 4.3). In order to simplify notation and avoid using the concept of principal angles, the above bound is relaxed to  $\text{aff}_{\mathcal{X}}^2$  in (11), as is accurate when two subspaces are orthogonal to each other. The relaxation leads to concise expression by only using affinity without aware of principal angles. Therefore, we recommend to apply the two bounds in respective situations.*

Finally, we want to highlight that the increment of affinity, the estimate accuracy, and the probability of deviation are determined by the dimensions of two subspaces, the original affinity, and the new ambient dimension, rather than the dimension of the original ambient space. This is also obvious.

Using Lemma 3 and Theorem 2, we may reach the concentration of distance between two subspaces after random projection.

**Corollary 2** *Let  $D_{\mathcal{X}}$  denote the distance between subspaces  $\mathcal{X}_1$  and  $\mathcal{X}_2$  of dimensions  $d_1 \leq d_2$ . The distance after Gaussian random projection,  $D_{\mathcal{Y}}$ , can be estimated by*

$$\overline{D}_{\mathcal{Y}}^2 = D_{\mathcal{X}}^2 - \frac{d_2}{n} \left( D_{\mathcal{X}}^2 - \frac{d_2 - d_1}{2} \right). \quad (13)$$

When  $n$  is large enough, the estimation error is bounded by

$$\mathbb{P} \left( \left| D_{\mathcal{Y}}^2 - \overline{D}_{\mathcal{Y}}^2 \right| > D_{\mathcal{X}}^2 \varepsilon \right) \leq \frac{4d_1}{\varepsilon^2 n}. \quad (14)$$

PROOF The proof is postponed to Section 4.3. ■

### 3.3 Restricted Isometry Property of random projection for subspaces

Based on the above results, we are ready to state the RIP of subspaces. We will begin with a special case of two given subspaces and then extend to a general case of any two candidates in a finite set of subspaces. As a consequence, we generalize the JL lemma from set of points to set of subspaces.

**Theorem 3** Suppose  $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^N$  are two subspaces with dimension  $d_1 \leq d_2$ , respectively. If  $\mathcal{X}_1$  and  $\mathcal{X}_2$  are projected into  $\mathbb{R}^n$  by a Gaussian random matrix  $\Phi \in \mathbb{R}^{n \times N}$ ,  $\mathcal{X}_k \xrightarrow{\Phi} \mathcal{Y}_k, k = 1, 2$ , then we have

$$(1 - \varepsilon)D_{\mathcal{X}}^2 \leq D_{\mathcal{Y}}^2 \leq (1 + \varepsilon)D_{\mathcal{X}}^2, \quad (15)$$

with probability at least

$$1 - \frac{4d_1}{(\varepsilon - d_2/n)^2 n}, \quad (16)$$

when  $n$  is large enough.

PROOF The proof is postponed to Section 4.4. ■

Theorem 3 shows that when  $n$  is sufficiently large, the distance between two subspaces remains unchanged with full probability after random projection.

According to Theorem 3, we can readily conclude the RIP of finite subspaces set in Theorem 1.

**Remark 7** It should be noticed that the RIP for all low dimensional subspaces doesn't hold in a way similar to sparse signals, even if for subspaces of dimension one.

**Remark 8** One may already notice that affinity, which measures similarity, can not provide a characteristic like RIP. An direct example is that for two independent subspaces with zero affinity, the projected affinity always deviates from zero. On the contrary, for two subspaces of zero distance, which only happens in the situation that the subspaces are exactly identical, the projected distance is of course zero. This visualizes our motivation of introducing the distance and building a metric space in Section 2.2.

## 4 Proofs of the Main Results

Before proving the main results, we would like to define Gaussian random vector and introduce an important tool in Lemma 6 for simplifying our conclusion.

**Definition 6** A Gaussian random vector  $\mathbf{a} \in \mathbb{R}^n$  has i.i.d. zero-mean Gaussian entries with variance  $\sigma^2 = 1/n$ .

**Lemma 6** Assume that  $X(n) \geq 0$  is a random variable indexed by  $n \in \mathbb{N}^+$  and for all positive  $\varepsilon_1, \varepsilon_2$ , it holds that

$$\mathbb{P}(X(n) > c_1 \varepsilon_1 + c_2 \varepsilon_2 | n) \leq p(\varepsilon_1, n) + q(\varepsilon_2, n), \quad (17)$$

where  $c_1, c_2$  are two positive constants, and  $p(\cdot), q(\cdot)$  are two real-valued functions. If

$$\lim_{n \rightarrow \infty} \frac{q(\varepsilon_2, n)}{p(\varepsilon_1, n)} = 0, \quad \forall \varepsilon_1, \varepsilon_2 > 0, \quad (18)$$

then we may simplify (17) by

$$\mathbb{P}(X(n) > c_1 \varepsilon |n) \leq p(\varepsilon, n), \quad (19)$$

when  $n$  is large enough.

PROOF The proof is postponed to Appendix 7.2. ■

#### 4.1 Proof of Lemma 4

We will first check some properties of Gaussian random vectors, which will be used in the proof.

**Lemma 7** *Let  $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$  are two Gaussian random vectors, which are dependent to each other. If  $\mathbb{E}\mathbf{p}\mathbf{q}^T = \omega \mathbf{I}_n/n, 0 \leq \omega \leq 1$ , we have*

$$\mathbb{P}\left(\left|\frac{\|\mathbf{p}\|^2}{\|\mathbf{q}\|^2} - 1\right| > (1 - \omega^2) \varepsilon\right) \leq \frac{4}{\varepsilon^2 n} =: P_1(\varepsilon, n), \quad (20)$$

when  $n$  is large enough.

PROOF The proof is postponed to Appendix 7.3. ■

**Lemma 8** *Let  $\mathbf{u} = \mathbf{a}/\|\mathbf{a}\|$ , where  $\mathbf{a} \in \mathbb{R}^n$  is a Gaussian random vectors. Let  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_d] \in \mathbb{R}^{n \times d}$  denote a given orthonormal matrix and  $\theta_i$  denote the angle between  $\mathbf{u}$  and  $\mathbf{v}_i, 1 \leq i \leq d$ , then we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^d \cos^2 \theta_i - \frac{d}{n}\right| > \varepsilon\right) < \frac{2d}{\varepsilon^2 n^2} =: P_2(\varepsilon, n). \quad (21)$$

PROOF The proof is postponed to Appendix 7.4. ■

Let us begin the proof of Lemma 4 by choosing the bases for the line  $\mathcal{X}_1$  and the subspace  $\mathcal{X}_2$  and then calculate the compressed affinity. One may refer to Figure 2 for a visualization.

According to the definition of affinity,  $\lambda = \cos \theta$ , where  $\theta$  is the only principal angle between  $\mathcal{X}_1$  and  $\mathcal{X}_2$ . We use  $\mathbf{u}$  and  $\mathbf{u}_1$  to denote the basis of  $\mathcal{X}_1$  and the unit vector, which constructs the principal angle with  $\mathbf{u}$ . Notice that  $\mathbf{u}_1$  locates inside  $\mathcal{X}_2$ . Consequently, we may decompose  $\mathbf{u}$  by

$$\mathbf{u} = \lambda \mathbf{u}_1 + \sqrt{1 - \lambda^2} \mathbf{u}_0,$$

where  $\mathbf{u}_0$  denotes some unit vector orthogonal to  $\mathcal{X}_2$ . Based on the above definition, we choose  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_d]$  as the basis of  $\mathcal{X}_2$ . Notice that  $\{\mathbf{u}_2, \dots, \mathbf{u}_d\}$  could be freely chosen when the orthogonality is satisfied.

After random projection, the basis of  $\mathcal{Y}_1$  changes to

$$\begin{aligned} \mathbf{a} &= \Phi \mathbf{u} = \lambda \Phi \mathbf{u}_1 + \sqrt{1 - \lambda^2} \Phi \mathbf{u}_0 \\ &= \lambda \mathbf{a}_1 + \sqrt{1 - \lambda^2} \mathbf{a}_0, \end{aligned} \quad (22)$$

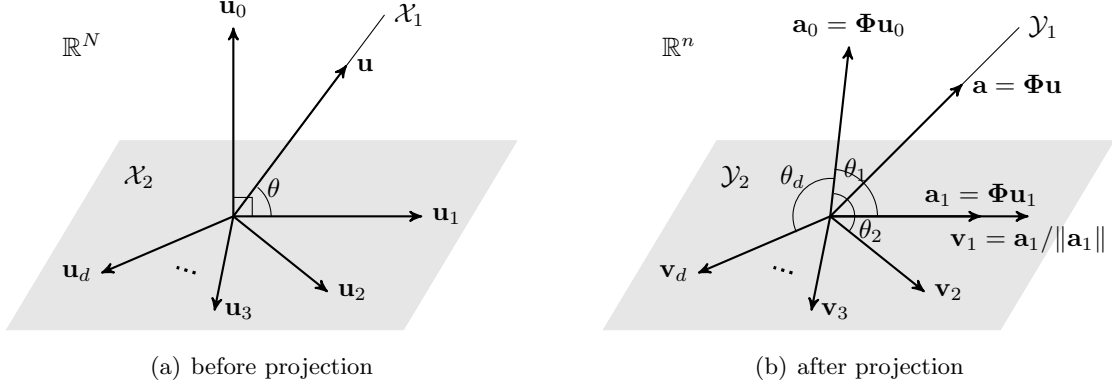


Figure 2: A geometric explanation of Lemma 4.

where  $\mathbf{a}_1 = \Phi \mathbf{u}_1$  and  $\mathbf{a}_0 = \Phi \mathbf{u}_0$  are not orthogonal to each other. As to  $\mathcal{Y}_2$ , considering that  $\Phi \mathbf{U}$  is not an orthonormal basis, we do orthogonalization by using Gram-Schmidt process. Denote the orthogonalized matrix by  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_d]$ , the first column of which

$$\mathbf{v}_1 = \mathbf{a}_1 / \|\mathbf{a}_1\| \quad (23)$$

does not change its direction after the orthogonalization.

According to the definition of affinity in (3), we may calculate the compressed affinity by

$$\text{aff}_{\mathcal{Y}}^2 = \left\| \frac{\mathbf{a}^T \mathbf{V}}{\|\mathbf{a}\|} \right\|^2 = \frac{1}{\|\mathbf{a}\|^2} \sum_{i=1}^d (\mathbf{a}^T \mathbf{v}_i)^2 \quad (24)$$

Splitting the summation in (24) into two parts and using (22), (23), and that  $\mathbf{V}$  is an orthonormal matrix, we have

$$\begin{aligned} \text{aff}_{\mathcal{Y}}^2 &= \frac{1}{\|\mathbf{a}\|^2} \left( (\lambda \mathbf{a}_1^T \mathbf{v}_1 + \sqrt{1-\lambda^2} \mathbf{a}_0^T \mathbf{v}_1)^2 + \sum_{i=2}^d (\sqrt{1-\lambda^2} \mathbf{a}_0^T \mathbf{v}_i)^2 \right) \\ &= \frac{1}{\|\mathbf{a}\|^2} \left( \lambda^2 \|\mathbf{a}_1\|^2 + 2\lambda \sqrt{1-\lambda^2} \|\mathbf{a}_0\| \|\mathbf{a}_1\| \cos \theta_1 + \sum_{i=1}^d (1-\lambda^2) \|\mathbf{a}_0\|^2 \cos^2 \theta_i \right), \end{aligned} \quad (25)$$

where  $\theta_i$  denote the angles between  $\mathbf{a}_0$  and  $\mathbf{v}_i$  for  $i = 1, \dots, d$ . By taking the norm on both sides of (22), we write

$$\begin{aligned} \|\mathbf{a}\|^2 &= \left\| \lambda \mathbf{a}_1 + \sqrt{1-\lambda^2} \mathbf{a}_0 \right\|^2 \\ &= \lambda^2 \|\mathbf{a}_1\|^2 + 2\lambda \sqrt{1-\lambda^2} \|\mathbf{a}_0\| \|\mathbf{a}_1\| \cos \theta_1 + (1-\lambda^2) \|\mathbf{a}_0\|^2. \end{aligned} \quad (26)$$

Eliminating  $\|\mathbf{a}_1\|$  by inserting (26) into (25), we get

$$\begin{aligned} \text{aff}_{\mathcal{Y}}^2 &= \frac{1}{\|\mathbf{a}\|^2} \left( \|\mathbf{a}\|^2 - (1-\lambda^2) \|\mathbf{a}_0\|^2 + \sum_{i=1}^d (1-\lambda^2) \|\mathbf{a}_0\|^2 \cos^2 \theta_i \right) \\ &= 1 - (1-\lambda^2) \frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2} \left( 1 - \sum_{i=1}^d \cos^2 \theta_i \right). \end{aligned} \quad (27)$$

We are ready for estimating  $\|\mathbf{a}_0\|^2/\|\mathbf{a}\|^2$  and  $\sum_{i=1}^d \cos^2 \theta_i$  by using Lemma 7 and Lemma 8, respectively. First recalling Lemma 7, let  $\mathbf{p} = \mathbf{a}_0$  and  $\mathbf{q} = \mathbf{a}$ . Using (22) we have

$$\mathbb{E}\mathbf{a}_0\mathbf{a}^\top = \sqrt{1-\lambda^2}\mathbb{E}\mathbf{a}_0\mathbf{a}_0^\top = \sqrt{1-\lambda^2}\mathbf{I}_n/n.$$

Denoting  $\omega = \sqrt{1-\lambda^2}$  and applying Lemma 7, we have

$$\mathbb{P}\left(\left|\frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2} - 1\right| > \lambda^2\varepsilon_1\right) \leq \frac{4}{\varepsilon_1^2 n} = P_1(\varepsilon_1, n). \quad (28)$$

Then recalling Lemma 8 and that  $\mathbf{a}_0$  is independent with  $\mathbf{V}$  by using the properties of Gaussian random distribution, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^d \cos^2 \theta_i - \frac{d}{n}\right| > \varepsilon_2\right) < \frac{2d}{\varepsilon_2^2 n^2} = P_2(\varepsilon_2, n). \quad (29)$$

Consequently, combing (27) and (6), the estimate error is rewritten as

$$\begin{aligned} \left|\text{aff}_y^2 - \overline{\text{aff}}_y^2\right| &= \left|\text{aff}_y^2 - \left(\lambda^2 + \frac{d}{n}(1-\lambda^2)\right)\right| \\ &= \left|1 - (1-\lambda^2)\frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2}\left(1 - \sum_{i=1}^d \cos^2 \theta_i\right) - \left(1 - (1-\lambda^2)\left(1 - \frac{d}{n}\right)\right)\right| \\ &= (1-\lambda^2)\left|\left(1 - \frac{d}{n}\right) - \frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2}\left(1 - \sum_{i=1}^d \cos^2 \theta_i\right)\right|. \end{aligned} \quad (30)$$

By using (28) and (29), the second item in RHS of (30) is bounded by

$$\begin{aligned} \left|\left(1 - \frac{d}{n}\right) - \frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2}\left(1 - \sum_{i=1}^d \cos^2 \theta_i\right)\right| &\leq \left|\left(1 - \frac{d}{n}\right) - (1 + \lambda^2\varepsilon_1)\left(\left(1 - \frac{d}{n}\right) + \varepsilon_2\right)\right| \\ &\leq \left(1 - \frac{d}{n}\right)\lambda^2\varepsilon_1 + \varepsilon_2 + \lambda^2\varepsilon_1\varepsilon_2 \end{aligned} \quad (31)$$

$$\sim \left(1 - \frac{d}{n}\right)\lambda^2\varepsilon_1 + \varepsilon_2, \quad (32)$$

with probability at least  $1 - P_1(\varepsilon_1, n) - P_2(\varepsilon_2, n)$ , where the last item in (31) is dropped.

Next we will use Lemma 6 to simplify (32). Recalling that  $d$  is much smaller than  $n$ , we have

$$\lim_{n \rightarrow \infty} \frac{P_2(\varepsilon_2, n)}{P_1(\varepsilon_1, n)} = \frac{d\varepsilon_1^2}{n\varepsilon_2^2} = 0.$$

Finally, inserting (32) in (30) and using Lemma 6, we have

$$\mathbb{P}\left(\left|\text{aff}_y^2 - \overline{\text{aff}}_y^2\right| > (1-\lambda^2)\left(1 - \frac{d}{n}\right)\lambda^2\varepsilon\right) \leq P_1(\varepsilon, n) = \frac{4}{\varepsilon^2 n}. \quad (33)$$

Using  $d \ll n$  again to drop  $(1 - d/n)$  from (33), we complete the proof.

## 4.2 Proof of Lemma 5 and Theorem 2

Before the detailed proof, we briefly introduce two major techniques utilized here.

First, we introduce a *quasi*-orthonormal basis of the original lower dimensional subspace for estimating the projected affinity. In order to calculate the affinity, recalling its definition in (3), we need to prepare the orthonormal basis for both subspaces. According to our recent study [36], however, a reasonable way is to utilize the normalized data matrix to approximate its Gram-Schmidt orthogonalization. Let  $\mathbf{U}_k$  denote the orthonormal basis of  $\mathcal{X}_k$ . After projection, the basis matrix becomes  $\mathbf{A}_k = \Phi \mathbf{U}_k$ , whose columns are not orthogonal to each other. We may process  $\mathbf{A}_k$  by using Gram-Schmidt process to yield the accurate orthogonal basis  $\mathbf{V}_k$ . Or, according to the Corollary 1 in [36] (referred as Lemma 10 in this work), we may normalize each column of  $\mathbf{A}_k$  to produce  $\bar{\mathbf{A}}_k$  as a rather good estimate of  $\mathbf{V}_k$ . Then  $\bar{\mathbf{A}}_1^T \mathbf{V}_2$  can be used as  $\mathbf{V}_1^T \mathbf{V}_2$  to estimate the projected affinity. Second, we deem the original lower dimensional subspace as  $d_1$  independent one-dimensional subspaces and then calculate the distance between  $\bar{\mathbf{A}}_1^T \mathbf{V}_2$  and the estimator (10) by using Lemma 4.

Now we need to introduce more properties of Gaussian random matrix to fulfill the proof.

**Lemma 9** [12, 13, 14] *Let  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$  be two independent Gaussian random vectors. Let  $\theta$  denote the angle between  $\mathbf{a}_1$  and  $\mathbf{a}_2$ , then we have*

$$\mathbb{P}(|\cos \theta| > \varepsilon) \leq \exp\left(-\frac{\varepsilon^2 n}{2}\right) =: P_3(\varepsilon, n). \quad (34)$$

PROOF Equation (34) is verified by using the concentration of measure. ■

**Lemma 10** [36] *Let  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \in \mathbb{R}^{n \times d}$  denote the Gram-Schmidt orthogonalization of a column-normalized matrix  $\bar{\mathbf{A}} = [\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \dots, \bar{\mathbf{a}}_d]$ , where  $\|\bar{\mathbf{a}}_i\| = 1, \forall i$ . Denote*

$$\bar{\mathbf{R}} = (\bar{r}_{ji}) = \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \mathbf{I}.$$

*When  $\bar{r}_{ji} = \bar{\mathbf{a}}_j^T \bar{\mathbf{a}}_i$  is small enough for  $j \neq i$ , we can use  $\bar{\mathbf{A}}$  to approximate  $\mathbf{V}$ . For an arbitrary vector  $\mathbf{w} \in \mathbb{R}^n$ , we conclude that*

$$\left| \|\mathbf{V}^T \mathbf{w}\|^2 - \|\bar{\mathbf{A}}^T \mathbf{w}\|^2 \right| \leq d \|\bar{\mathbf{A}}^T \mathbf{w}\|^2 \max \bar{\mathbf{R}} + \epsilon(\bar{\mathbf{R}}) \|\bar{\mathbf{R}}\|_F, \quad (35)$$

*where  $\lim_{\bar{\mathbf{R}} \rightarrow 0} \epsilon(\bar{\mathbf{R}}) = 0$ .*

**Corollary 3** *Follow the definition of Lemma 10 and assume  $\bar{r}_{ji} < \varepsilon, \forall j \neq i$ . When  $\varepsilon$  is small enough, for an arbitrary matrix  $\mathbf{W} \in \mathbb{R}^{n \times l}$ , we conclude that*

$$\left| \|\mathbf{V}^T \mathbf{W}\|_F^2 - \|\bar{\mathbf{A}}^T \mathbf{W}\|_F^2 \right| \leq d \|\bar{\mathbf{A}}^T \mathbf{W}\|_F^2 \varepsilon.$$

PROOF Denote  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_l] \in \mathbb{R}^{n \times l}$ . According to Lemma 10, when  $\varepsilon$  is small enough, we have

$$\begin{aligned} \left| \|\mathbf{V}^T \mathbf{W}\|_F^2 - \|\bar{\mathbf{A}}^T \mathbf{W}\|_F^2 \right| &\leq \sum_{i=1}^l \left| \|\mathbf{V}^T \mathbf{w}_i\|_2^2 - \|\bar{\mathbf{A}}^T \mathbf{w}_i\|_2^2 \right| \\ &\leq d \sum_{i=1}^l \|\bar{\mathbf{A}}^T \mathbf{w}_i\|_2^2 \varepsilon \\ &= d \|\bar{\mathbf{A}}^T \mathbf{W}\|_F^2 \varepsilon. \quad \blacksquare \end{aligned}$$

Before going into the proof of Lemma 5 and Theorem 2, we first simplify the notation. for any matrix  $(\cdot)_k$  or its column vector  $(\cdot)_{k,i}$  in this subsection, the subscript  $k$  denotes the index of subspaces, i.e.,  $k = 1, 2$ .

There are four steps in this proof. We will prepare the basis matrices for the subspaces before and after projection in the first step. Then following a proof sketch, Lemma 5 will be justified in the last three steps. Finally Theorem 2 is reached by relaxing some conditions.

**Step 1)** Let  $\tilde{\mathbf{U}}_k = [\tilde{\mathbf{u}}_{k,1}, \dots, \tilde{\mathbf{u}}_{k,d_k}]$  denote any orthonormal matrix for subspace  $\mathcal{X}_k$ . According to the definition of affinity in (3), one may do singular value decomposition to  $\tilde{\mathbf{U}}_2^T \tilde{\mathbf{U}}_1$ ,

$$\tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_2 = \mathbf{Q}_1 \mathbf{\Lambda} \mathbf{Q}_2^T,$$

where  $\mathbf{Q}_1$  and  $\mathbf{Q}_2^T$  denote, respectively, the orthonormal basis of the column space and row space for  $\tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_2$ . The singular values  $\lambda_i = \cos \theta_i, 1 \leq i \leq d_1$  is located on the diagonal of  $\mathbf{\Lambda}$ , where  $\theta_i$  denotes the  $i$ th principal angle. After reshaping, we have

$$\left( \tilde{\mathbf{U}}_1 \mathbf{Q}_1 \right)^T \tilde{\mathbf{U}}_2 \mathbf{Q}_2 = \mathbf{U}_1^T \mathbf{U}_2 = \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \left| \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \right. \\ & \ddots & & \\ & & \lambda_{d_1} & \left| \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \right. \end{bmatrix},$$

where  $\mathbf{U}_k = \tilde{\mathbf{U}}_k \mathbf{Q}_k$  are the orthonormal basis, which has the closest connection with the affinity between these two subspaces. Specifically, the angles of the first  $d_1$  columns are all principal angles for calculating affinity, i.e.

$$\mathbf{u}_{1,i}^T \mathbf{u}_{2,j} = \begin{cases} \lambda_i, & 1 \leq i = j \leq d_1; \\ 0 & \text{elsewhere.} \end{cases} \quad (36)$$

After projection by using Gaussian random matrix  $\Phi$ , the original basis matrix changes to  $\mathbf{A}_k = \Phi \mathbf{U}_k = [\mathbf{a}_{k,1}, \dots, \mathbf{a}_{k,d_k}]$ , whose columns are not orthogonal to each other. Considering the angles between any two columns are not very large, however, we may normalize each columns as

$$\bar{\mathbf{A}}_k = [\bar{\mathbf{a}}_{k,1}, \dots, \bar{\mathbf{a}}_{k,d_k}] = \left[ \frac{\mathbf{a}_{k,1}}{\|\mathbf{a}_{k,1}\|}, \dots, \frac{\mathbf{a}_{k,d_k}}{\|\mathbf{a}_{k,d_k}\|} \right],$$



which could be used to approximate the orthonormal basis of  $\mathcal{Y}_k$ . In order to obtain the accurate orthonormal basis for the compressed subspace, the efficient method is to process  $\bar{\mathbf{A}}_k$  by using Gram-Schmidt orthogonalization. We use  $\mathbf{V}_k = [\mathbf{v}_{k,1}, \dots, \mathbf{v}_{k,d_k}]$  to denote the orthonormal basis after orthogonalization.

Now we are ready to introduce the sketch of our proof. In Step 2), we use  $\bar{\mathbf{A}}_1^T \mathbf{V}_2$  to estimate the compressed affinity according to Corollary 3. In Step 3), we first deem the original subspace  $\mathcal{X}_1$  as  $d_1$  independent one-dimensional subspaces and then calculate the distance between  $\bar{\mathbf{A}}_1^T \mathbf{V}_2$  and the estimator of (10) by using Lemma 4. Finally, we combine the results in the above two steps and simplify it to complete the proof in the last step.

**Step 2)** According to the properties of Gaussian matrix, we know that  $\bar{\mathbf{a}}_{1,i}$  and  $\bar{\mathbf{a}}_{1,j}$ , which are obtained by random projection and normalization, are independent for all  $i \neq j$ . Using Lemma 9 and as a consequence, with probability at least  $1 - (d_1(d_1 - 1)/2) P_3(\varepsilon_3, n)$ , we have

$$|\bar{\mathbf{a}}_{1,i}^T \bar{\mathbf{a}}_{1,j}| \leq \varepsilon_3, \quad \forall 1 \leq i \neq j \leq d_1.$$

This means that  $\bar{\mathbf{A}}_1$  well approximates  $\mathbf{V}_1$  and can be roughly utilized as an orthonormal basis. Recalling the definition of affinity in (3) and using Corollary 3, we have

$$\begin{aligned} \left| \text{aff}_{\mathcal{Y}}^2 - \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 \right| &= \left| \|\mathbf{V}_1^T \mathbf{V}_2\|_F^2 - \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 \right| \\ &\leq d_1 \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 \varepsilon_3. \end{aligned} \quad (37)$$

**Step 3)** Now we will deem all basis vectors of  $\mathcal{X}_1$  separately as multiple one-dimensional subspaces, denoted by  $\mathcal{X}_{1,i}, 1 \leq i \leq d_1$ . According to its definition and (36), the affinity between  $\mathcal{X}_{1,i}$  and  $\mathcal{X}_2$  equals  $\lambda_i$ . Actually we are interested in the relation between  $\mathcal{X}_{1,i}$  and  $\mathcal{X}_2$  after random projection. This has been solved by Lemma 4, which means that, with probability at least  $1 - P_1(\varepsilon_4, n)$ ,

$$\left| \text{aff}_{\mathcal{Y}_i}^2 - \overline{\text{aff}}_{\mathcal{Y}_i}^2 \right| \leq \lambda_i^2 (1 - \lambda_i^2) \varepsilon_4, \quad (38)$$

where

$$\begin{aligned} \text{aff}_{\mathcal{Y}_i}^2 &= \|\bar{\mathbf{a}}_{1,i}^T \mathbf{V}_2\|^2, \\ \overline{\text{aff}}_{\mathcal{Y}_i}^2 &= \lambda_i^2 + \frac{d_2}{n} (1 - \lambda_i^2), \end{aligned} \quad (39)$$

denote, respectively, the affinity and its estimate between the compressed line  $\mathcal{Y}_{1,i}$  and the compressed subspace  $\mathcal{Y}_2$ . Equation (39) comes from that  $\bar{\mathbf{a}}_{1,i}$  is the orthonormal basis for  $\mathcal{Y}_{1,i}$ .

Considering the independence among these one-dimensional subspaces and using (2),

(10), (38), and (39), we have, with probability at least  $1 - d_1 P_1(\varepsilon_4, n)$ ,

$$\begin{aligned}
\left| \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 - \overline{\text{aff}}_y^2 \right| &= \left| \sum_{i=1}^{d_1} \|\bar{\mathbf{a}}_{1,i}^T \mathbf{V}_2\|^2 - \left( \text{aff}_x^2 + \frac{d_2}{n} (d_1 - \text{aff}_x^2) \right) \right| \\
&= \left| \sum_{i=1}^{d_1} \text{aff}_{y_i}^2 - \sum_{i=1}^{d_1} \left( \lambda_i^2 + \frac{d_2}{n} (1 - \lambda_i^2) \right) \right| \\
&\leq \sum_{i=1}^{d_1} \left| \text{aff}_{y_i}^2 - \overline{\text{aff}}_{y_i}^2 \right| \\
&\leq \sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \varepsilon_4. \tag{40}
\end{aligned}$$

**Step 4)** Combining (37) and (40) by utilizing triangle inequality, we readily reach that, with probability at least  $1 - (d_1(d_1 - 1)/2) P_3(\varepsilon_3, n) - d_1 P_1(\varepsilon_4, n)$ ,

$$\begin{aligned}
\left| \text{aff}_y^2 - \overline{\text{aff}}_y^2 \right| &\leq \left| \text{aff}_y^2 - \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 \right| + \left| \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 - \overline{\text{aff}}_y^2 \right| \\
&\leq d_1 \|\bar{\mathbf{A}}_1^T \mathbf{V}_2\|_F^2 \varepsilon_3 + \sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \varepsilon_4.
\end{aligned}$$

Recalling Lemma 6 and that  $P_3(\varepsilon_3, n)$  decreases exponentially with respect to  $n$ , we have

$$\begin{aligned}
\mathbb{P} \left( \left| \text{aff}_y^2 - \overline{\text{aff}}_y^2 \right| > \sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \varepsilon \right) &\leq d_1 P_1(\varepsilon_4, n) + \frac{d_1(d_1 - 1)}{2} P_3(\varepsilon_3, n) \tag{41} \\
&\sim d_1 P_1(\varepsilon, n) = \frac{4d_1}{\varepsilon^2 n},
\end{aligned}$$

We then complete the proof of Lemma 5. Finally, relaxing the bound in (41) by

$$\sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \leq \sum_{i=1}^{d_1} \lambda_i^2 = \text{aff}_x^2,$$

Theorem 2 is proved.

### 4.3 Proof of Corollary 2

By reshaping (13), we have

$$D_y^2 - \bar{D}_y^2 = (D_y^2 - D_x^2) + \frac{d_2}{n} \left( D_x^2 - \frac{d_2 - d_1}{2} \right). \tag{42}$$

Using Lemma 3 in (42), we are ready to verify

$$D_y^2 - \bar{D}_y^2 = \overline{\text{aff}}_y^2 - \text{aff}_y^2. \tag{43}$$

Now let check the bound in (12),

$$\sum_{i=1}^{d_1} \lambda_i^2 (1 - \lambda_i^2) \leq \sum_{i=1}^{d_1} (1 - \lambda_i^2) = d_1 - \text{aff}_x^2 \leq D_x^2. \tag{44}$$

Combing (43) and (44) in Lemma 5, the proof is complete.

#### 4.4 Proof of Theorem 3

Jointly applying the fact of  $D_{\mathcal{X}}^2 \geq (d_2 - d_1)/2$  by Lemma 3 and triangle inequality in (42), we have

$$|D_{\mathcal{Y}}^2 - D_{\mathcal{X}}^2| \leq |D_{\mathcal{Y}}^2 - \overline{D_{\mathcal{Y}}^2}| + \frac{d_2}{n} \left( D_{\mathcal{X}}^2 - \frac{d_2 - d_1}{2} \right). \quad (45)$$

Inserting (45) in (14), we have

$$\mathbb{P} \left( |D_{\mathcal{Y}}^2 - D_{\mathcal{X}}^2| > D_{\mathcal{X}}^2 \varepsilon_1 + \frac{d_2}{n} \left( D_{\mathcal{X}}^2 - \frac{d_2 - d_1}{2} \right) \right) \leq \frac{4d_1}{\varepsilon_1^2 n}. \quad (46)$$

By redefining  $\varepsilon$  as

$$D_{\mathcal{X}}^2 \varepsilon = D_{\mathcal{X}}^2 \varepsilon_1 + \frac{d_2}{n} \left( D_{\mathcal{X}}^2 - \frac{d_2 - d_1}{2} \right), \quad (47)$$

we have

$$\frac{4d_1}{\varepsilon_1^2 n} = \frac{4d_1}{\left( \varepsilon - \frac{d_2}{n} + \frac{d_2(d_2 - d_1)}{2nD_{\mathcal{X}}^2} \right)^2 n} < \frac{4d_1}{\left( \varepsilon - \frac{d_2}{n} \right)^2 n}. \quad (48)$$

Using (47) and (48) in (46), the proof is complete.

## 5 Numerical verification

In this section, the main result of Theorem 2 is evaluated by numerical simulations. In order to save computation, we randomly generate two subspaces in the following steps.

1. Given  $d_1 \leq d_2 \ll N$ , generate an orthonormal matrix  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{d_1+d_2}] \in \mathbb{R}^{N \times (d_1+d_2)}$ .
2. Let  $\mathbf{U}_2 = [\mathbf{w}_1, \dots, \mathbf{w}_{d_2}]$  be the orthonormal basis for subspace  $\mathcal{X}_2$ .
3. Given affinity  $\text{aff}_{\mathcal{X}}$ , randomly choose  $\hat{\lambda}_i, 1 \leq i \leq d_1$  from the uniform distribution in  $[0, 1]$  and then scale them to the affinity, i.e.

$$\lambda_i = \text{aff}_{\mathcal{X}} \cdot \frac{\hat{\lambda}_i}{\left( \sum_{i=1}^{d_1} \hat{\lambda}_i^2 \right)^{\frac{1}{2}}}.$$

4. Calculate the orthonormal basis for subspace  $\mathcal{X}_1$  as

$$\mathbf{U}_1 = \left[ \lambda_1 \mathbf{w}_1 + (1 - \lambda_1^2)^{\frac{1}{2}} \mathbf{w}_{d_2+1}, \dots, \lambda_{d_1} \mathbf{w}_{d_1} + (1 - \lambda_{d_1}^2)^{\frac{1}{2}} \mathbf{w}_{d_2+d_1} \right].$$

With this method, we can generate two subspaces with any given affinity, which are ready for projection.

In the first experiment, the estimate of the compressed affinity (10) is verified in the condition of  $(N, n) = (500, 200)$  and  $(d_1, d_2) = (5, 10)$ . The original affinity in the ambient space is chosen as  $\text{aff}_{\mathcal{X}}^2 = 1, 2, 3, 4$ , respectively. For each  $\text{aff}_{\mathcal{X}}^2$ , a random Gaussian matrix

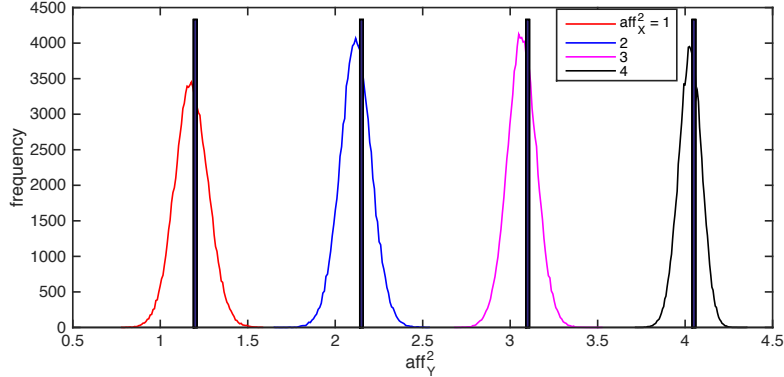


Figure 3: This figure demonstrates the experimental frequency (denoted by curves) and the theoretical estimate (denoted by bars) of the compressed affinity, where  $(N, n) = (500, 200)$ ,  $(d_1, d_2) = (5, 10)$ , and the original affinities are fixed as 1, 2, 3, and 4. The frequencies are calculated by  $1E5$  trials.

is generated and used to project the two subspaces into the compressed space, where the compressed affinity  $\text{aff}_y^2$  is calculated. The frequencies of the compressed affinities obtained from  $1E5$  trials as well as with their theoretical estimates are demonstrated in Figure 3. One may read that the proposed estimate is rather accurate and the compressed affinities concentrate on their theoretical estimates.

In the second experiment, the estimate of the compressed affinity is further tested for all possible original affinities and by various subspace dimension combinations, where the dimensions of the ambient space and compressed space are the same as those in the first experiment. Here  $(d_1, d_2)$  is chosen from a candidate set and the original affinity varies from 0 to its maximum, i.e.,  $d_1 \leq d_2$ . For each case, two original subspaces and a random Gaussian matrix are generated, then the compressed affinity is calculated after projection. After repeating 500 times, the frequencies at different compressed affinities are computed and normalized by its maximum, i.e., the compressed affinity with the highest appearance is assigned 1 and the others are smaller than 1. Then the normalized frequencies for all cases are plotted in Figure 4, where the blue line denotes the theoretical estimate. This result further verifies that the compressed affinities of various dimensions of subspaces display the concentration property, as shows in Theorem 2.

The third experiment tests the effect of  $N$  and  $n$  in Theorem 2. By fixing  $(d_1, d_2) = (5, 10)$ , the compressed affinity of two subspaces being projected from an  $N$ -dimension space to an  $n$ -dimension space, where  $(N, n)$  is chosen from a candidate set, is shown. The result is plotted in Figure 5 by using the same way as that in the second experiment. One may readily find that by increasing  $n$ , the compressed affinity demonstrates better concentration. Whereas the dimension of the original space,  $N$ , has no effect on the concentration behavior. The observation agrees with Theorem 2.

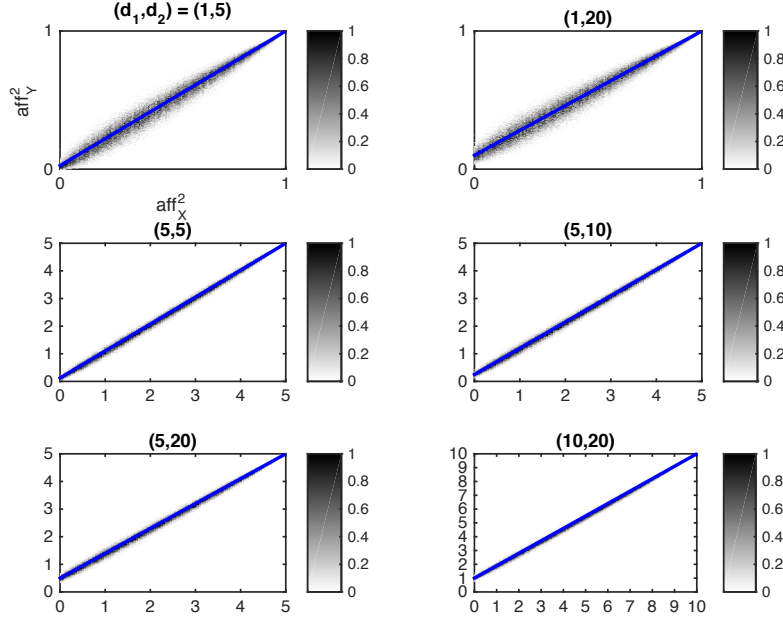


Figure 4: This figure demonstrates the experimental compressed affinity (which frequency is denoted by gray area) and the theoretical estimate (denoted by blue line), where  $(N, n) = (500, 200)$  and  $(d_1, d_2)$  are displayed on the title.

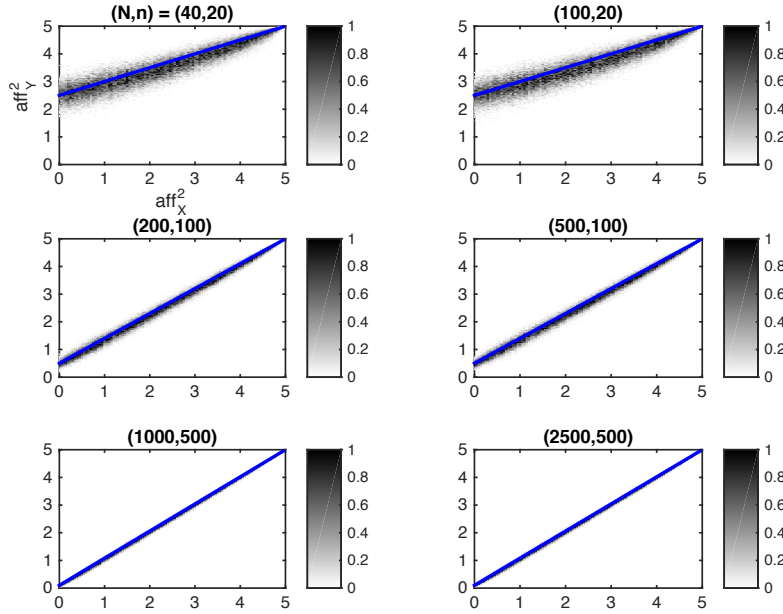


Figure 5: This figure demonstrates the experimental compressed affinity (which frequency is denoted by gray area) and the theoretical estimate (denoted by blue line), where  $(d_1, d_2) = (5, 10)$  and  $(N, n)$  are displayed on the title.

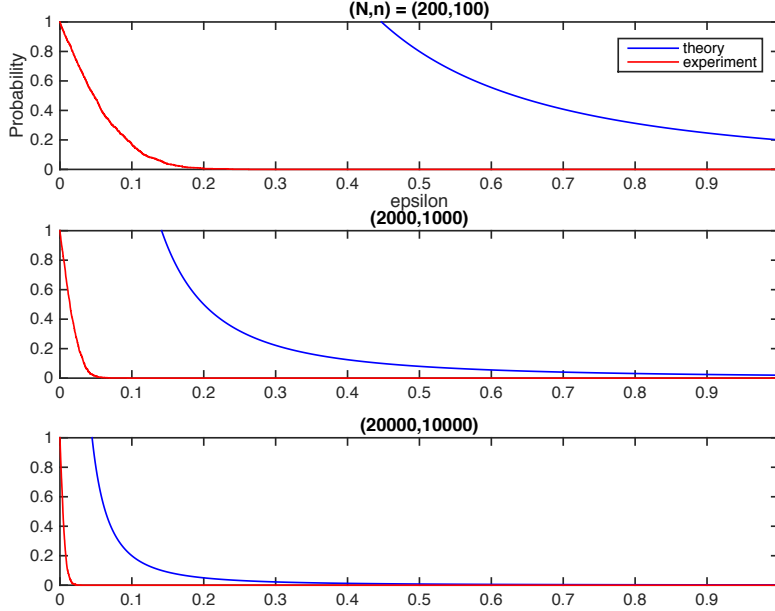


Figure 6: This figure demonstrates the experimental error (denoted by red curve) and its upper bound (denoted by blue curve) of the estimated compressed affinity, where  $(d_1, d_2) = (5, 10)$ , the original affinities are fixed as 2, and  $(N, n)$  are displayed on the title.

In the last experiment, the upper bound of the estimated compressed affinity (11) in Theorem 2 is verified numerically. By fixing  $(d_1, d_2) = (5, 10)$ ,  $\text{aff}_{\mathcal{X}}^2 = 2$  and choosing  $(N, n)$  from a candidate set, the probability that the estimate error falls out of the bound is plotted in Figure 6, where the blue line denotes the theoretical result of (11), and the red line denotes the experimental result of 1E3 trials. One may read that as  $n$  increases the theoretical bound approaches to the experimental result gradually. This verifies that Theorem 2 is rigid.

## 6 Conclusion

This work generates the JL Lemma and the RIP from finite signal set and sparse signals, respectively, to subspaces. We study the distance-preserving property of Gaussian random projection for subspaces. By introducing the projection  $F$ -norm distance and build a metric space, the connection between affinity and distance are revealed for the first time. We then theoretically prove that with high probability the affinity or distance between two projected subspaces are concentrated on their estimates. When the new ambient dimension is sufficiently large, the affinity and distance between two subspaces almost remain unchanged after random projection. Finally, the main contribution is generalized to a finite set of subspaces. Numerical experiments verify the theoretical work.

## 7 Appendix

### 7.1 Proof of Lemma 3

Denote the orthonormal basis of subspace  $\mathcal{X}_k$  by  $\mathbf{U}_k = [\mathbf{u}_{k,1}, \dots, \mathbf{u}_{k,d_k}]$ ,  $k = 1, 2$ . According to the definition of distance in (4), we have

$$\begin{aligned} D^2(\mathcal{X}_1, \mathcal{X}_2) &= \frac{1}{2} \|\mathbf{P}_1 - \mathbf{P}_2\|_F^2 \\ &= \frac{1}{2} \|\mathbf{U}_1 \mathbf{U}_1^\top - \mathbf{U}_2 \mathbf{U}_2^\top\|_F^2 \\ &= \frac{1}{2} \left\| \sum_{i=1}^{d_1} \mathbf{u}_{1,i} \mathbf{u}_{1,i}^\top - \sum_{j=1}^{d_2} \mathbf{u}_{2,j} \mathbf{u}_{2,j}^\top \right\|_F^2. \end{aligned} \quad (49)$$

Denote the  $l$ th entry of  $\mathbf{u}_{k,i}$  by  $u_{k,li}$  and expand the RHS of (49), we have

$$\begin{aligned} D^2(\mathcal{X}_1, \mathcal{X}_2) &= \frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n \left( \sum_{i=1}^{d_1} u_{1,li} u_{1,mi} - \sum_{j=1}^{d_2} u_{2,lj} u_{2,mj} \right)^2 \\ &=: \frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n (A_1 + A_2 - 2B), \end{aligned} \quad (50)$$

where

$$\begin{aligned} A_k &= \left( \sum_{i=1}^{d_k} u_{k,li} u_{k,mi} \right)^2 \\ &= \sum_{i=1}^{d_k} u_{k,li}^2 u_{k,mi}^2 + \sum_{1 \leq i \neq j \leq d_k} u_{k,li} u_{k,mi} u_{k,lj} u_{k,mj}, \end{aligned} \quad (51)$$

$$B = \sum_{i=1}^{d_1} u_{1,li} u_{1,mi} \sum_{j=1}^{d_2} u_{2,lj} u_{2,mj}. \quad (52)$$

We will study the three items separately. First, using (51) in (50) and changing the order of summation, we have

$$\begin{aligned} \sum_{l=1}^n \sum_{m=1}^n A_k &= \sum_{i=1}^{d_k} \left( \sum_{l=1}^n u_{k,li}^2 \sum_{m=1}^n u_{k,mi}^2 \right) + \sum_{1 \leq i \neq j \leq d_k} \left( \sum_{l=1}^n u_{k,li} u_{k,lj} \sum_{m=1}^n u_{k,mi} u_{k,mj} \right) \\ &= \sum_{i=1}^{d_k} \|\mathbf{u}_{k,i}\|^2 \|\mathbf{u}_{k,i}\|^2 + \sum_{1 \leq i \neq j \leq d_k} \mathbf{u}_{k,i}^\top \mathbf{u}_{k,j} \mathbf{u}_{k,i}^\top \mathbf{u}_{k,j}. \end{aligned}$$

Considering that  $\mathbf{u}_{k,i}$ ,  $1 \leq i \leq d_k$  are columns drawn from an orthonormal matrix, we have

$$\sum_{l=1}^n \sum_{m=1}^n A_k = \sum_{i=1}^{d_k} 1 + \sum_{1 \leq i \neq j \leq d_k} 0 = d_k. \quad (53)$$

Now we check the last item in (50). Using (52) in (50) and changing the order of summation, we have

$$\begin{aligned} \sum_{l=1}^n \sum_{m=1}^n B &= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \left( \sum_{l=1}^n u_{1,li} u_{2,lj} \sum_{m=1}^n u_{1,mi} u_{2,mj} \right) \\ &= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} |\mathbf{u}_{1,i}^T \mathbf{u}_{2,j}|^2. \end{aligned}$$

Recalling the definition of affinity in (3), we have

$$\sum_{l=1}^n \sum_{m=1}^n B = \text{aff}^2(\mathcal{X}_1, \mathcal{X}_2). \quad (54)$$

We then complete the proof by inserting (53) and (54) in (50).

## 7.2 Proof of Lemma 6

By introducing  $\varepsilon = \varepsilon_1 + c_2 \varepsilon_2 / c_1$ , we have  $\varepsilon_1 = (1 - 1/m)\varepsilon$ , and  $\varepsilon_2 = c_1 \varepsilon / (m c_2)$  for all  $m \in \mathbb{N}^+$ . Using (18) in (17), we have, for large  $n$ ,

$$\begin{aligned} \mathbb{P}(X(n) > c_1 \varepsilon | n) &\leq p((1 - 1/m)\varepsilon, n) + q\left(\frac{c_1}{m c_2} \varepsilon, n\right) \\ &\sim p(\varepsilon(1 - 1/m), n). \end{aligned}$$

Let  $m$  approach infinity, and then (19) is obtained.

## 7.3 Proof of Lemma 7

First, we introduce Lemma 11, from which Lemma 7 is extended.

**Lemma 11** *Let  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$  are two independent Gaussian random vectors. Since  $\|\mathbf{a}_1\|^2 \sim \chi_n^2/n$  and  $\|\mathbf{a}_2\|^2 \sim \chi_n^2/n$  are independent,*

$$F_{n,n} \triangleq \frac{\|\mathbf{a}_1\|^2}{\|\mathbf{a}_2\|^2}$$

*follows an F-distribution with parameters  $(n, n)$ , so we have*

$$\mathbb{P}(|F_{n,n} - 1| > \varepsilon) \leq \frac{4}{\varepsilon^2 n} = P_1(\varepsilon, n), \quad (55)$$

*when  $n$  is large enough.*

PROOF According to the properties of F-distribution, we have

$$\begin{aligned} \mathbb{E}F_{n,n} &= \frac{n}{n-2}, \\ \text{Var}(F_{n,n}) &= \frac{4n(n-1)}{(n-2)^2(n-4)}. \end{aligned}$$



Then (55) is verified by using Chebyshev's inequality as

$$\begin{aligned}\mathbb{P}(|F_{n,n} - 1| > \varepsilon) &\leq \frac{\mathbb{E}|F_{n,n} - 1|^2}{\varepsilon^2} \\ &= \frac{\text{Var}(F_{n,n}) + (\mathbb{E}F_{n,n} - 1)^2}{\varepsilon^2} \\ &= \frac{4(n+2)}{\varepsilon^2(n-2)(n-4)} \sim \frac{4}{\varepsilon^2 n}.\end{aligned}\quad \blacksquare$$

Now we begin the proof of Lemma 7 by introducing an assistant vector

$$\mathbf{w} = \frac{\mathbf{p} - \omega \mathbf{q}}{\sqrt{1 - \omega^2}}, \quad (56)$$

which is orthogonal to  $\mathbf{q}$ . This can be verified by

$$\begin{aligned}\mathbb{E}\mathbf{w}\mathbf{q}^T &= \frac{1}{\sqrt{1 - \omega^2}} (\mathbb{E}\mathbf{p}\mathbf{q}^T - \omega \mathbb{E}\mathbf{q}\mathbf{q}^T) \\ &= \frac{1}{\sqrt{1 - \omega^2}} \left( \frac{\omega}{n} \mathbf{I}_n - \frac{\omega}{n} \mathbf{I}_n \right) = \mathbf{0}.\end{aligned}\quad (57)$$

Using the above orthogonality and the given condition, we further write

$$\begin{aligned}\mathbb{E}\mathbf{w}\mathbf{w}^T &= \frac{1}{\sqrt{1 - \omega^2}} \mathbb{E}\mathbf{w}\mathbf{p}^T \\ &= \frac{1}{1 - \omega^2} (\mathbb{E}\mathbf{p}\mathbf{p}^T - \omega \mathbb{E}\mathbf{q}\mathbf{p}^T) \\ &= \frac{1}{1 - \omega^2} \left( \frac{1}{n} \mathbf{I}_n - \frac{\omega^2}{n} \mathbf{I}_n \right) = \frac{1}{n} \mathbf{I}_n.\end{aligned}$$

These show that  $\mathbf{q}$  and  $\mathbf{w}$  are independent Gaussian random vectors. Following Lemma 11, we denote  $\|\mathbf{w}\|^2/\|\mathbf{q}\|^2$  by  $F_{n,n}$ .

By representing  $\mathbf{p}$  with  $\mathbf{w}$  using (56), consequently, we have

$$\begin{aligned}\frac{\|\mathbf{p}\|^2}{\|\mathbf{q}\|^2} &= \frac{\|\omega \mathbf{q} + \sqrt{1 - \omega^2} \mathbf{w}\|^2}{\|\mathbf{q}\|^2} \\ &= \omega^2 + (1 - \omega^2) \frac{\|\mathbf{w}\|^2}{\|\mathbf{q}\|^2} + 2\omega\sqrt{1 - \omega^2} \frac{\|\mathbf{w}\|}{\|\mathbf{q}\|} \cos \theta \\ &= \omega^2 + (1 - \omega^2) F_{n,n} + 2\omega\sqrt{1 - \omega^2} \sqrt{F_{n,n}} \cos \theta,\end{aligned}\quad (58)$$

where  $\theta$  is the angle between  $\mathbf{q}$  and  $\mathbf{w}$ . Recalling Lemma 11 and Lemma 9, we have, with probability at least  $1 - P_1(\varepsilon_5, n) - P_3(\varepsilon_6, n)$ ,

$$\left| (1 - \omega^2)(F_{n,n} - 1) + 2\omega\sqrt{1 - \omega^2} \sqrt{F_{n,n}} \cos \theta \right| \leq (1 - \omega^2)\varepsilon_5 + 2\omega\sqrt{1 - \omega^2} \sqrt{1 + \varepsilon_5} \varepsilon_6, \quad (59)$$

$$\sim (1 - \omega^2)\varepsilon_5 + 2\omega\sqrt{1 - \omega^2} \varepsilon_6, \quad (60)$$

where  $\sqrt{1 + \varepsilon_5}$  in (59) comes from (55) and it is then approximated by one in (60) because  $\varepsilon_5$  is a small quantity.

Now we will use Lemma 6 to simplify (60). Because  $P_3(\varepsilon_6, n)$  decreases exponentially with respect to  $n$  and the condition of Lemma 6 is satisfied, we have, when  $n$  is large,

$$\begin{aligned} \mathbb{P} \left( \left| (1-\omega^2)(F_{n,n}-1) + 2\omega\sqrt{1-\omega^2}\sqrt{F_{n,n}}\cos\theta \right| > (1-\omega^2)\varepsilon \right) &\leq P_1(\varepsilon_5, n) + P_3(\varepsilon_6, n) \\ &\sim P_1(\varepsilon, n) = \frac{4}{\varepsilon^2 n}. \end{aligned} \quad (61)$$

Combining (58) and (61), the proof is completed.

## 7.4 Proof of Lemma 8

Lemma 8 is a corollary of the following Lemma.

**Lemma 12** *Let  $\mathbf{w} = \mathbf{a}/\|\mathbf{a}\|$ , where  $\mathbf{a} \in \mathbb{R}^n$  is a Gaussian random vector. For any support  $\mathcal{T} \subset [1 : n]$  with cardinality  $d \triangleq |\mathcal{T}|$ , we have*

$$\mathbb{P} \left( \left| \|\mathbf{w}_{\mathcal{T}}\|^2 - \frac{d}{n} \right| > \varepsilon \right) < \frac{2d}{\varepsilon^2 n^2} = P_2(\varepsilon, n), \quad (62)$$

where  $\mathbf{w}_{\mathcal{T}}$  is composed by the entries of  $\mathbf{w}$  supported on  $\mathcal{T}$ .

PROOF By calculating, we have

$$\begin{aligned} \mathbb{E}\|\mathbf{w}_{\mathcal{T}}\|^2 &= \frac{d}{n}, \\ \text{Var}(\|\mathbf{w}_{\mathcal{T}}\|^2) &= \frac{2d(n-d)}{n^2(n+2)} \end{aligned}$$

Then (62) is verified by using Chebyshev's inequality as

$$\mathbb{P} \left( \left| \|\mathbf{w}_{\mathcal{T}}\|^2 - \frac{d}{n} \right| > \varepsilon \right) \leq \frac{2d(n-d)}{\varepsilon^2 n^2(n+2)} < \frac{2d}{\varepsilon^2 n^2}. \quad \blacksquare$$

Recalling the definition of  $\mathbf{w}$  in Lemma 12, we may let  $w_i = \cos \phi_i$ , where  $\phi_i$  denotes the angle between  $\mathbf{w}$  and the  $i$ th coordinate axis,  $\mathbf{e}_i$ . Because the relation of  $\mathbf{w}$  with respect to  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_d]$  and that of  $\mathbf{u}$  with respect to  $\mathbf{V}$  are exactly identical, (21) can be readily verified.

## References

- [1] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz maps into a hilbert space," vol. 26, no. 189, pp. 189–206, 1984.
- [2] S. Dasgupta and A. Gupta, "An elementary proof of the johnson-lindenstrauss lemma," *International Computer Science Institute, Technical Report*, pp. 99–006, 1999.
- [3] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

- [4] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathématique*, vol. 346, no. 9–10, pp. 589–592, 2008.
- [5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, no. 28, pp. 253–263, 2015.
- [6] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [8] S. Aeron, V. Saligrama, and M. Zhao, “Information theoretic bounds for compressed sensing,” *Information Theory IEEE Transactions on*, vol. 56, no. 10, pp. 5111–5130, 2010.
- [9] E. Candes and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse problems*, vol. 23, no. 3, p. 969, 2007.
- [10] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.
- [11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “The johnson-lindenstrauss lemma meets compressed sensing,” *Submitted manuscript, June*, 2006.
- [12] P. Lévy and F. Pellegrino, *Problèmes concrets d’analyse fonctionnelle: avec un complément sur les fonctionnelles analytiques*. Gauthier-Villars, 1951.
- [13] D. Achlioptas, “Database-friendly random projections,” in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2001, pp. 274–281.
- [14] M. Ledoux, *The concentration of measure phenomenon*. American Mathematical Soc., 2005, no. 89.
- [15] J. Haupt and R. Nowak, “A generalized restricted isometry property,” *University of Wisconsin*, vol. 189, no. 6, pp. S99–S99, 2007.
- [16] L. H. Chang and J. Y. Wu, “Achievable angles between two compressed sparse vectors under norm/distance constraints imposed by the restricted isometry property: A plane geometry approach,” *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2059 – 2081, 2012.

- [17] K. Gedalyahu and Y. C. Eldar, “Time-delay estimation from low-rate samples: A union of subspaces approach,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3017–3031, 2010.
- [18] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [19] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [20] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [21] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [22] M. F. Duarte and Y. C. Eldar, “Structured compressed sensing: From theory to applications,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.
- [23] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, “Signal processing with compressive measurements,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.
- [24] T. Blumensath and M. E. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *Information Theory IEEE Transactions on*, vol. 55, no. 4, pp. 1872–1882, 2009.
- [25] A. Eftekhari and M. B. Wakin, “New analysis of manifold embeddings and signal recovery from compressive measurements,” *Applied And Computational Harmonic Analysis*, vol. 39, no. 1, pp. 67–109, 2013.
- [26] R. G. Baraniuk and M. B. Wakin, “Random projections of smooth manifolds,” *Foundations of Computational Mathematics*, vol. 9, no. 1, pp. 51–77, 2010.
- [27] H. L. Yap, M. B. Wakin, and C. J. Rozell, “Stable manifold embeddings with structured random matrices,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 4, pp. 720–730, 2013.
- [28] E. Elhamifar and R. Vidal, “Sparse subspace clustering,” 2009, pp. 2790–2797.

- [29] M. Soltanolkotabi, E. J. Candes *et al.*, “A geometric analysis of subspace clustering with outliers,” *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.
- [30] 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.
- [31] R. Heckel and H. Bölcskei, “Robust subspace clustering via thresholding,” *Information Theory IEEE Transactions on*, vol. 61, no. 11, pp. 6320–6342, 2015.
- [32] R. Heckel, M. Tschannen, and H. Bolcskei, “Subspace clustering of dimensionality-reduced data,” in *Information Theory (ISIT), 2014 IEEE International Symposium on*, 2014, pp. 2997–3001.
- [33] X. Mao and Y. Gu, “Compressed subspace clustering: A case study,” in *Signal and Information Processing*, 2014, pp. 453 – 457.
- [34] R. Heckel, M. Tschannen, and H. Bölcskei, “Dimensionality-reduced subspace clustering,” *arXiv preprint arXiv:1507.07105*, 2015.
- [35] Y. Wang, Y.-X. Wang, and A. Singh, “A theoretical analysis of noisy sparse subspace clustering on dimensionality-reduced data,” *arXiv preprint arXiv:1610.07650*, 2016.
- [36] G. Li and Y. Gu, “Approximation of gram-schmidt orthogonalization by data matrix,” *arXiv preprint arXiv:1701.00711*, 2016.
- [37] J. Hamm and D. D. Lee, “Grassman discriminant analysis: a unifying view on subspace-based learning,” in *International Conference*, 2008, pp. 376–383.
- [38] J. Miao and A. Ben-Israel, “On principal angles between subspaces in  $\mathbb{R}^n$ ,” *Linear Algebra And Its Applications*, vol. 171, no. 92, pp. 81–98, 1992.
- [39] L. Qiu, Y. Zhang, and C. K. Li, “Unitarily invariant metrics on the grassmann space,” *Siam Journal on Matrix Analysis And Applications*, vol. 27, no. 2, pp. 507–531, 2005.
- [40] C. Jordan, “Essai sur la géométrie à  $n$  dimensions,” *Bulletin de la Société mathématique de France*, vol. 3, pp. 103–174, 1875.
- [41] A. Galántai and H. C. J., “Jordan’s principal angles in complex vector spaces,” *NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS*, vol. 13, pp. 589–598, 2006.
- [42] A. Björck and G. H. Golub, “Numerical methods for computing the angles between linear subspaces,” *Mathematics of Computation*, vol. 27, pp. 579–594, 1973.
- [43] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

- [44] P.-A. Absil, R. Mahony, and R. Sepulchre, “Riemannian geometry of grassmann manifolds with a view on algorithmic computation,” *Acta Applicandae Mathematicae*, vol. 80, no. 2, pp. 199–220, 2004.