DISTANCE-PRESERVING PROPERTY OF RANDOM PROJECTION FOR SUBSPACES

Gen Li and Yuantao Gu

Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

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 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, we study for the first time the distance-preserving property of Gaussian random projection matrices for two subspaces based on knowledge of Grassmann manifold. We theoretically prove that with high probability the *affinity* or the *distance* between two compressed subspaces are concentrated on their estimates. Numerical experiments verify the theoretical work.

Index Terms— Johnson-Lindenstrauss Lemma, Restricted Isometry Property, Gaussian matrix, Union of Subspaces, affinity

1. INTRODUCTION

In a big data era, the dimensionality of problems increases rapidly. Dimension reduction is important to reduce the complexity of solving methods. For example, Restricted Isometry Property (RIP) [1, 2, 3] admits the use of random projection to reduce the dimension while keeping the Euclidean distance, which leads to the boom of Compressed Sensing (CS) and sparsity related researches [4, 5, 6]. In CS, it has been shown that, with high probability, a random projection of a high-dimensional but sparse or compressible signal onto a low-dimensional space can be recovered robustly due to the RIP of some random projections. 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 vector, $\mathbf{y} \in \mathbb{R}^n (n < N)$ is the compressed vector, and $\mathbf{\Phi} \in \mathbb{R}^{n\times N}$ is a random projection matrix. To sufficiently ensure unique representation and robust recovery to the original signal, the random projection matrix should approximately preserve the distance between any two k-sparse signals. Specifically, the well-known Johnson-Lindenstrauss (JL) Lemma [7] states that, with high probability, there exists a constant $0 < \varepsilon < 1$, such that

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

This work was partially supported by National Natural Science Foundation of China (NSFC 61371137, 61571263, 61531166005, 51459003) and Tsinghua University Initiative Scientific Research Program (Grant 2014Z01005). The corresponding author of this work is Yuantao Gu (gyt@tsinghua.edu.cn).

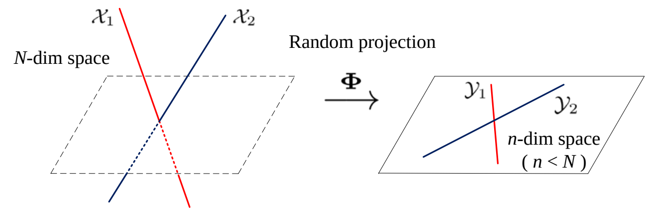

Fig. 1. This paper studies the distance-preserving property of Gaussian random projection matrices for two subspaces.

And RIP is a generalization of this lemma. In addition, there are theoretical results showing some angle-preserving properties as well [8, 9].

Furthermore, in [10, 11, 12], 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 [12, 13, 14]. It has been proved in [15, 16] that, with high probability the random projection matrix $\mathbf{\Phi}$ can preserve the length of a signal as well as the distance between two signals that lie in an UoS. Recently, the stable embedding property has been extended to signals modeled as low-dimensional Riemannian sub-manifolds in Euclidean space [17, 18, 19].

Very recently, there have been many researches focusing on different aspects of Subspace Clustering (SC) [20, 21, 22]. A ready approach to reduce the complexity of SC is to compress the original samples into low dimensional vectors and then to cluster the subspaces in the low dimensional space, which is called Compressed Subspace Clustering [23, 24]. The affinity between two subspaces determines their separability.

1.1. Main contribution

Although the UoS model is the most popular signal model and is extensively used in various applications, few theoretical analysis describes the embedding performance on these linear subspaces via random measurement matrices. In this paper, motivated by JL Lemma we study the distance-preserving property of Gaussian random projection matrices for two subspaces based on knowledge of Grassmann manifold. First, we use the fact that subspaces are points on the Grassmann manifold to define a metric on the set. Then, we prove the JL Lemma for two subspaces, that is, with high probability there exists a constant $0 < \varepsilon < 1$, such that

$$(1-\varepsilon)d^2(\mathcal{X}_1, \mathcal{X}_2) \leq d^2(\mathcal{Y}_1, \mathcal{Y}_2) \leq (1+\varepsilon)d^2(\mathcal{X}_1, \mathcal{X}_2),$$

where $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}_1, \mathcal{Y}_2$ are the original subspaces and compressed subspaces, respectively. Finally, we find the relationship between the distance and affinity of two subspaces, and draw conclusion to the affinity to provide a theoretical guarantee for CSC.

Although different metrics and distance measures have been used to describe the topological structure of the Grassmann manifold [25, 26, 27], as far as we know, there is no rigorous theoretical analysis for the distance-preserving property of subspaces. This paper theoretically studies this problem for the first time [1].

2. PROBLEM AND MOTIVATION

The space consisted of all subspaces of $\mathbb{R}^n$ is denoted by $\bigcup_{r=0}^{n} \mathbf{Gr}(r,n)$, where $\mathbf{Gr}(r,n)$ denotes a Grassmann manifold. Because each subspace corresponds to one and only one projection matrix, a distance measure on the space is defined by

$$d(\mathcal{X}_1,\mathcal{X}_2) = \frac{1}{\sqrt{2}}\|\mathbf{P}_{\mathcal{X}_1} - \mathbf{P}_{\mathcal{X}_2}\|_F,$$

where $\mathcal{X}_1,\mathcal{X}_2$ are two subspaces of $\mathbb{R}^n$ and $\mathbf{P}_{\mathcal{X}_1},\mathbf{P}_{\mathcal{X}_2}$ are the corresponding projection matrices, respectively. One may readily find that this definition meets all requirements in the definition of distance measure, thus the space becomes a metric space. In addition, if we define the affinity between two subspaces as

$$\mathrm{aff}(\mathcal{X}_1,\mathcal{X}_2) = \|\mathbf{U}_1^{\mathrm{T}}\mathbf{U}_2\|_F,$$

where $\mathbf{U}_1,\mathbf{U}_2$ are, respectively, orthonormal matrices of $\mathcal{X}_1,\mathcal{X}_2$ with dimension d_1 and d_2, we will find that there is a close relationship between the distance and the affinity.

Lemma 1 *The distance and affinity between two subspaces are related by*

$$d^2(\mathcal{X}_1,\mathcal{X}_2) = \frac{d_1+d_2}{2} - \mathrm{aff}^2(\mathcal{X}_1,\mathcal{X}_2).$$

Suppose there are two subspaces $\mathcal{X}_1,\mathcal{X}_2 \subset \mathbb{R}^N$ with dimensions, respectively, $d_1,d_2 \ll N$. The projection matrix $\mathbf{\Phi} \in \mathbb{R}^{n\times N}$, $n<N$, is composed of entries independently drawn from Gaussian distribution $\mathcal{N}(0,1/n)$. The dimension-reduced data that randomly projected by $\mathbf{\Phi}$ compose an n-dimension ambient space and the original subspaces change to

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

Assuming $d_1 \le d_2 \le n$, the dimension of subspaces remains unchanged after random projection in statistical sense. When $d_1 = 1$, $\mathcal{X}_1$ reduces to a vector.

In this paper, we will study the separability of subspaces after random projection. We use the *affinity* and the *distance* to measure the separability between two subspaces before and after compression, respectively, by

$$\mathrm{aff}_{\mathcal{X}} = \mathrm{aff}(\mathcal{X}_1,\mathcal{X}_2), \qquad d_{\mathcal{X}} = d(\mathcal{X}_1,\mathcal{X}_2),$$

and

$$\mathrm{aff}_{\mathcal{Y}} = \mathrm{aff}(\mathcal{Y}_1,\mathcal{Y}_2), \qquad d_{\mathcal{Y}} = d(\mathcal{Y}_1,\mathcal{Y}_2).$$

According to its definition, we can also write

$$\mathrm{aff}_{\mathcal{X}} = \left(\sum_{i=1}^{d_1}\lambda_i^2\right)^{\frac{1}{2}},$$

where λ_i is the absolute singular values of $\mathbf{U}_1^{\mathrm{T}}\mathbf{U}_2$. Or we can define λ_i equivalently as

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

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

3. MAIN RESULTS

In this section, we present our results on the distance-preserving property of subspaces after random projection. It begins with a simple case of estimating the compressed affinity of a vector and a subspace, and then extends the result to the case of two subspaces. Finally, the Restricted Affinity Property is stated.

Before introducing the main results, we would like to emphasize that the notation of *less than* in this work holds in the sense of equivalence. For example, if $f(n) \le \frac{1}{n-2}$, we may state that, without confusion, $f(n) \le \frac{1}{n}$ when n is large enough for simplicity, considering that $\frac{1}{n-2} \sim \frac{1}{n}$.

3.1. Estimated affinity between compressed subspace and vector

We first study the separability of a vector and a subspace after random compression.

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

$$\overline{\mathrm{aff}}^2 = \lambda^2 + \frac{d}{n}\left(1-\lambda^2\right), \tag{1}$$

where the estimation error is controlled by

$$\mathbb{P}\left(|\mathrm{aff}_{\mathcal{Y}}^2 - \overline{\mathrm{aff}}^2| > \lambda^2(1-\lambda^2)\varepsilon\right) \le \frac{4}{\varepsilon^2 n}, \tag{2}$$

when n is large enough.

PROOF According to the assumption and the definition of affinity, we can write the unit basis of $\mathcal{X}_1$ as $\mathbf{u} = \lambda\mathbf{u}_1 + \sqrt{1-\lambda^2}\mathbf{u}_0$ where $\mathbf{u}_1$ is some unit vector in the subspace $\mathcal{X}_2$ and $\mathbf{u}_0$ is some unit vector which is orthogonal to $\mathcal{X}_2$. Then we can choose $\mathbf{U} = (\mathbf{u}_1,...,\mathbf{u}_d)$ as an orthonormal matrix of $\mathcal{X}_2$. After the random projection, let $\mathbf{V}$ denote the orthonormal matrix of $\mathbf{\Phi}\mathbf{U}$ transformed by Gram-Schmidt process and $\mathbf{a} = \mathbf{\Phi}\mathbf{u} = \lambda\mathbf{a}_1 + \sqrt{1-\lambda^2}\mathbf{a}_0$. Since $\mathbf{a}_1$ has the same direction with $\mathbf{v}_1$, we have

$$\mathrm{aff}_{\mathcal{Y}}^2 = 1 - (1-\lambda^2)\frac{\|\mathbf{a}_0\|^2}{\|\mathbf{a}\|^2}\left(1 - \sum_{i=1}^{d}\cos^2\theta_i\right),$$

where θ_i denote the angles between $\mathbf{a}_0$ and $\mathbf{v}_i$ for $i=1,...,d$. Finally we can estimate $\|\mathbf{a}_0\|^2/\|\mathbf{a}\|^2$ and $\sum_{i=1}^{d}\cos^2\theta_i$ separately to get the result. Please refer to [28] for details. ∎

Using Lemma 1 and Lemma 2, we may readily reach the distance-preserving property of a vector and a subspace. Let $d_{\mathcal{X}} = \sqrt{(d+1)/2-\lambda^2}$ denote the distance between vector $\mathcal{X}_1$ and d-dimension subspace $\mathcal{X}_2$, then the distance after projection, $d_{\mathcal{Y}}$, can be estimated by

[1] The full version of this work is available at [28], which includes the detailed proofs and remarks.

$$\overline{d}^2 = d_{\mathcal{X}}^2 - \frac{d}{n}\left(d_{\mathcal{X}}^2 - \frac{d-1}{2}\right).$$

When n is large enough, the estimation error is controlled by

$$\mathbb{P}\left(|d_{\mathcal{Y}}^2 - \overline{d}^2| > \lambda^2(1-\lambda^2)\varepsilon\right) \leq \frac{4}{\varepsilon^2 n}. \quad (3)$$

3.2. Estimated affinity between two compressed subspaces

We then study the separability of two subspaces after random compression by similar mathematical tools as the ones used in Lemma 2.

Theorem 1 *Suppose $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^N$ are two subspaces with dimension d_1, d_2, respectively. Let $d_m = \min\{d_1, d_2\}$ and $d_M = \max\{d_1, d_2\}$. Define*

$$\overline{\mathrm{aff}}^2 = \mathrm{aff}_{\mathcal{X}}^2 + \frac{d_M}{n}(d_m - \mathrm{aff}_{\mathcal{X}}^2) \quad (4)$$

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

$$\mathbb{P}\left(|\mathrm{aff}_{\mathcal{Y}}^2 - \overline{\mathrm{aff}}^2| > \mathrm{aff}_{\mathcal{X}}^2\varepsilon\right) \leq \frac{4d_m}{\varepsilon^2 n}. \quad (5)$$

PROOF Assume that $d_2 \leq d_1$ and $\mathrm{aff}_{\mathcal{X}}^2 = \sum_{i=1}^{d_2}\lambda_i^2$. We can choose $\mathbf{U}_k = (\mathbf{u}_{k,1}, ..., \mathbf{u}_{k,d_k})$ as the orthonormal matrix of $\mathcal{X}_k$ for $k = 1, 2$, respectively, such that

$$\begin{aligned}\mathbf{U}_2 &= (\mathbf{u}_{2,1}, ..., \mathbf{u}_{2,d_2}) \\ &= \left(\lambda_1\mathbf{u}_{1,1} + \sqrt{1-\lambda_1^2}\mathbf{u}_1, ..., \lambda_{d_2}\mathbf{u}_{2,d_2} + \sqrt{1-\lambda_{d_2}^2}\mathbf{u}_{d_2}\right), \quad (6)\end{aligned}$$

where $\mathbf{u}_{1,1}, \cdots, \mathbf{u}_{1,d_2}$ and $\mathbf{u}_1, \cdots, \mathbf{u}_{d_2}$ are orthogonal. Let $\mathbf{V}_k = (\mathbf{v}_{k,1}, ..., \mathbf{v}_{k,d_k})$ denote the orthonormal matrix of $\mathbf{A}_k = \mathbf{\Phi}\mathbf{U}_k = (\mathbf{a}_{k,1}, ..., \mathbf{a}_{k,d_k})$ for $k = 1, 2$. Let $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_{d_2}) = (\mathbf{a}_{2,1}/\|\mathbf{a}_{2,1}\|, ..., \mathbf{a}_{2,d_2}/\|\mathbf{a}_{2,d_2}\|)$. The sketch of proof goes as follows. We first use $\mathbf{A}^{\mathrm{T}}\mathbf{V}_1$ to estimate $\mathrm{aff}_{\mathcal{Y}}^2 = \mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1$ [29]. Then, with Lemma 2, we get the estimator $\overline{\mathrm{aff}}^2 = \mathrm{aff}_{\mathcal{X}}^2 + \frac{d_1}{n}(d_2 - \mathrm{aff}_{\mathcal{X}}^2)$ through estimating every row of $\mathbf{A}^{\mathrm{T}}\mathbf{V}_1$, that is the affinity between the row of $\mathbf{A}$ and $\mathbf{V}_1$. Finally, we simplify the result to achieve the conclusion. Please refer to [28] for details. ■

One may easily check that Theorem 1 agrees with Lemma 2 when $d_m = 1$. Using Lemma 1 and Theorem 1, we may reach the distance-preserving property of two subspaces. Let $d_{\mathcal{X}} = \sqrt{(d_1+d_2)/2 - \mathrm{aff}_{\mathcal{X}}^2}$ denote the distance between $\mathcal{X}_1, \mathcal{X}_2$, then the distance after projection, $d_{\mathcal{Y}}$, can be estimated by

$$\overline{d}^2 = d_{\mathcal{X}}^2 - \frac{d_M}{n}\left(d_{\mathcal{X}}^2 - \frac{d_M - d_m}{2}\right).$$

When n is large enough, the estimation error is controlled by

$$\mathbb{P}\left(|d_{\mathcal{Y}}^2 - \overline{d}^2| > d_{\mathcal{X}}^2\varepsilon\right) \leq \frac{4d_m}{\varepsilon^2 n}. \quad (7)$$

3.3. Distance-preserving property of random projection for subspaces

Using triangle inequality in (7), we have

$$\mathbb{P}\left(|d_{\mathcal{Y}}^2 - d_{\mathcal{X}}^2| > \frac{d_M}{n}\left(d_{\mathcal{X}}^2 + \frac{d_m - d_M}{2}\right) + d_{\mathcal{X}}^2\varepsilon\right) \leq \frac{4d_m}{\varepsilon^2 n}. \quad (8)$$

By the fact that $\frac{d_M - d_m}{2} \leq d_{\mathcal{X}}^2 \leq \frac{d_M + d_m}{2}$, if we consider the error, $\frac{d_M}{n}\left(d_{\mathcal{X}}^2 + \frac{d_m - d_M}{2}\right) + d_{\mathcal{X}}^2\varepsilon$, in (8) as a whole, then we can conclude the following theorem.

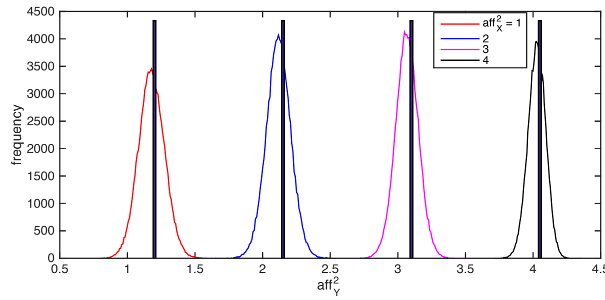

Fig. 2. This figure demonstrates the experimental frequency (denoted by curves) and the theoretical estimate (denoted by bars) of the compressed affinity, where $(N, n) = (500, 200)$, $(d_1, d_2) = (5, 10)$, and the original affinities are fixed as $1, 2, 3$, and 4. The frequencies are calculated by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

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

$$(1-\varepsilon)d_{\mathcal{X}}^2 \leq d_{\mathcal{Y}}^2 \leq (1+\varepsilon)d_{\mathcal{X}}^2, \quad (9)$$

with probability at least $1 - \frac{4d_m}{(\varepsilon - d_M/n)^2 n}$, when n is large enough.

Theorem 2 shows that when n is sufficiently large, the distance between two subspaces remains unchanged with probability 1 after Gaussian random projection. When n is large enough, the change of affinity after projection is controlled by

$$\mathbb{P}(|\mathrm{aff}_{\mathcal{Y}}^2 - \mathrm{aff}_{\mathcal{X}}^2| > \varepsilon) \leq \frac{4d_m \mathrm{aff}_{\mathcal{X}}^4}{(\varepsilon - \frac{d_M d_m}{n})^2 n}. \quad (10)$$

And the change of distance after projection is controlled by

$$\mathbb{P}(|d_{\mathcal{Y}}^2 - d_{\mathcal{X}}^2| > \varepsilon) \leq \frac{4d_m d_{\mathcal{X}}^4}{(\varepsilon - \frac{d_M d_m}{n})^2 n}. \quad (11)$$

Please notice that we will not draw a conclusion for affinity similar to Theorem 2, because the affinity is not a distance.

4. NUMERICAL SIMULATION

In this section, the main result of Theorem 1 is evaluated by numerical simulations. In order to save computation, we first randomly generate a subspace and then generate a second subspace by giving affinity as (6), where the λs are randomly generated from the uniform distribution in $[0, 1]$ and then scaled to the affinity. By 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 (4) 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 $\mathrm{aff}_{\mathcal{X}}^2 = 1, 2, 3, 4$, respectively. For each $\mathrm{aff}_{\mathcal{X}}^2$, a random Gaussian matrix is generated and used to project the two subspaces into the compressed space, where the compressed affinity $\mathrm{aff}_{\mathcal{Y}}^2$ is calculated. The frequencies of the compressed affinities obtained from 1E5 trials as well as with their theoretical estimates are demonstrated in Fig. 2. One may read that the proposed estimate is rather

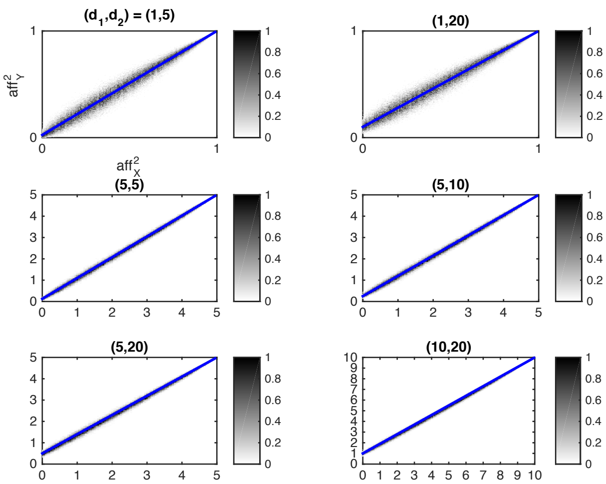

Fig. 3. This figure demonstrates the experimental compressed affinity (which frequency is denoted by gray area) and the theoretical estimate (denoted by blue line), where $(N, n) = (500, 200)$ and (d_1, d_2) are displayed on the title.

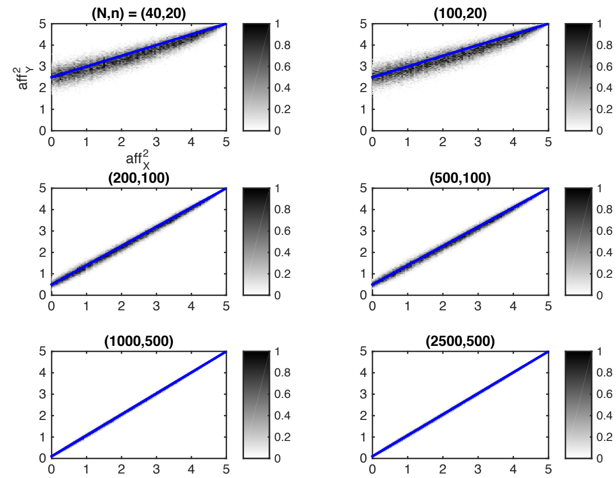

Fig. 4. This figure demonstrates the experimental compressed affinity (which frequency is denoted by gray area) and the theoretical estimate (denoted by blue line), where $(d_1, d_2) = (5, 10)$ and (N, n) are displayed on the title.

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 that 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., $\min(d_1, d_2)$. 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 Fig. 3, 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 1.

The third experiment tests the effect of N and n in Theorem 1. 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 Fig. 4 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 1.

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

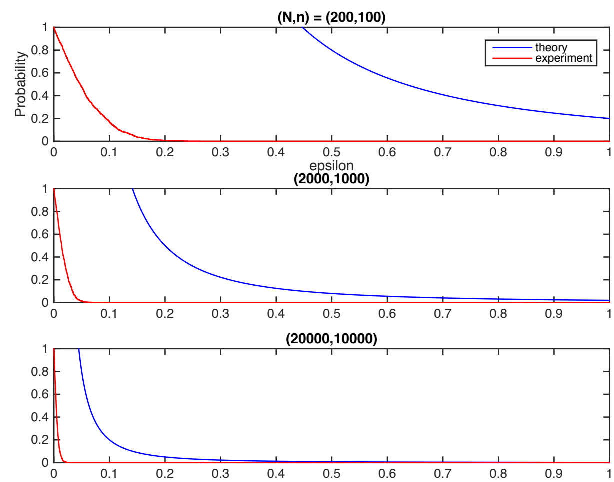

Fig. 5. This figure demonstrates the experimental error (denoted by red curve) and its upper bound (denoted by blue curve) of the estimated compressed affinity, where $(d_1, d_2) = (5, 10)$, the original affinities are fixed as 2, and (N, n) are displayed on the title.

5. CONCLUSION

In this paper, we formulated subspaces as points on Grassmann manifold, defined a metric for UoS, and found the relation between the definitions of distance and affinity, which is an important quantity for SC. Then we generalized JL Lemma in the case of compressing two subspaces and established the distance-preserving property. In addition, we provided numerical simulations for validation. However, how to prove the RIP for random projection for the space consisting of all low-dimensional subspaces is still an open problem.

6. REFERENCES

[1] Emmanuel J. Candes and Terence Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] Emmanuel J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9–10, pp. 589–592, 2008.

[3] Richard Baraniuk, Mark Davenport, Ronald Devore, and Michael Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 28, pp. 253–263, 2015.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] 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.

[6] S. Aeron, V. Saligrama, and Manqi Zhao, "Information theoretic bounds for compressed sensing," *Information Theory IEEE Transactions on*, vol. 56, no. 10, pp. 5111–5130, 2010.

[7] William B. Johnson and Joram Lindenstrauss, "Extensions of lipschitz maps into a hilbert space," vol. 26, no. 189, pp. 189–206, 1984.

[8] Jarvis Haupt and Robert Nowak, "A generalized restricted isometry property," *University of Wisconsin*, vol. 189, no. 6, pp. S99–S99, 2007.

[9] Ling Hua Chang and Jwo Yuh 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.

[10] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 41–47, 2008.

[11] 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.

[12] 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.

[13] S. F. Cotter, B. D. Rao, Kjersti 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.

[14] 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.

[15] 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.

[16] Thomas Blumensath and Mike 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.

[17] Armin Eftekhari and Michael 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.

[18] Richard G. Baraniuk and Michael B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, pp. 51–77, 2010.

[19] 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.

[20] E. Elhamifar and R. Vidal, "Sparse subspace clustering," 2009, pp. 2790–2797.

[21] Mahdi Soltanolkotabi, Ehsan Elhamifar, and Emmanuel J. Candès, "Robust subspace clustering," *Annals of Statistics*, vol. 42, no. 2, pp. 669–699, 2014.

[22] Reinhard Heckel and Helmut Bölcskei, "Robust subspace clustering via thresholding," *Information Theory IEEE Transactions on*, vol. 61, no. 11, pp. 6320–6342, 2015.

[23] 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.

[24] Xianghui Mao and Yuantao Gu, "Compressed subspace clustering: A case study," in *Signal and Information Processing*, 2014, pp. 453 – 457.

[25] Jihun Hamm and Daniel D. Lee, "Grassman discriminant analysis: a unifying view on subspace-based learning," in *International Conference*, 2008, pp. 376–383.

[26] Jianming Miao and Adi Ben-Israel, "On principal angles between subspaces in rn," *Linear Algebra And Its Applications*, vol. 171, no. 92, pp. 81–98, 1992.

[27] Li Qiu, Yanxia Zhang, and Chi Kwong Li, "Unitarily invariant metrics on the grassmann space," *Siam Journal on Matrix Analysis And Applications*, vol. 27, no. 2, pp. 507–531, 2005.

[28] Gen Li and Yuantao Gu, *Restricted Isometry Property (RIP) for Finite Set of Subspaces via Gaussian Random Projection*, preparing for submission, available at http://gu.ee.tsinghua.edu.cn/publications, Tsinghua University, Beijing, China, 2016.

[29] Gen Li and Yuantao Gu, *Approximation of Gram-Schmidt Orthogonalization by Data Matrix*, http://arxiv.org/abs/1701.00711.