Restricted Isometry Property of Subspace Projection Matrix Under Random Compression

Xinyue Shen, Student Member, IEEE, and Yuantao Gu, Member, IEEE

Abstract-Structures play a significant role in the field of signal processing. As a representative of structural data, low rank matrix along with its restricted isometry property (RIP) has been an important research topic in compressive signal processing. Subspace projection matrix is a kind of low rank matrix with additional structure, which allows for further reduction of its intrinsic dimension. This leaves room for improving its own RIP, which could work as the foundation of compressed subspace projection matrix recovery. In this work, we study the RIP of subspace projection matrix under random orthonormal compression. Considering the fact that subspace projection matrices of s dimensional subspaces in \mathbb{R}^N form an s(N-s) dimensional submanifold in $\mathbb{R}^{N \times N}$, our main concern is transformed to the stable embedding of such submanifold into $\mathbb{R}^{N \times N}$. The result is that by $O(s(N-s) \log N)$ number of random measurements the RIP of subspace projection matrix is guaranteed.

Index Terms—Compressive signal processing, low rank matrix, manifold stable embedding, restricted isometry property, subspace projection matrix.

I. INTRODUCTION

S IGNAL structure has always been a key point in the field of signal processing. Structural data, such as sparse signal and low rank matrix, have been important research topics in compressive signal processing [1], [2]. These structures invoke low intrinsic dimension, so the restricted isometry property (RIP) can be established to guarantee both exact and robust reconstructions from randomly compressed measurements [3], [4], [5].

For a given s dimensional linear subspace S in an Euclidean space \mathbb{R}^N , assuming that s is less than N, the subspace projection matrix is a low rank matrix with rather specific structure. In fact, it is not only symmetric, semi-definite, but also has merely eigenvalues 1 and 0. Such additional structure invokes lower intrinsic dimension than a general low rank matrix does, therefore theoretical improvement on the RIP can be expected.

According to the basic ideas in compressive sensing [6], the RIP of subspace projection matrix could work as the foundation of compressed subspace projection matrix recovery. Con-

Manuscript received December 08, 2014; revised February 04, 2015; accepted February 06, 2015. Date of publication February 10, 2015; date of current version February 16, 2015. This work was supported by National 973 Program of China under Grant 2013CB329203, and by the National Natural Science Foundation of China under Grant 61371137. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jie Liang.

The authors are with the Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China (e-mail: gyt@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2402206

sidering the fact that subspace projection matrix has a one to one correspondence with subspace, the recovery from its compression could be viewed as compressed subspace estimation. Subspace estimation has been a concerning problem in signal processing and computer vision. In some scenarios, such as face recognition [7], motion segmentation [8], and visual tracking [9], the objects belong to subspaces with much lower dimension than the ambient space. In fact, subspace estimation from highly incomplete information has recently appeared as an attractive research topic [10], [11], [12].

For subspaces with a given dimension in \mathbb{R}^N , their projection matrices form a manifold. There are significant and solid works in manifold-modeled signal recovery from randomly compressed measurements [13], [14], [15], [16]. These works extend classic compressed sensing by generalizing low-dimension model from sparse signal to signal on low dimensional manifold, and study stable manifold embeddings and nonadaptive dimensionality reduction of data on manifold. One of the key ideas is to control the regularity of the manifold so that it is well-conditioned. The work in [17] utilizes an instructive quantity called the condition number of a manifold also known as the reach of a manifold, which studies submanifold extrinsically and unveils its Riemannian geometry properties [18].

In this work, we aim to study the RIP of subspace projection matrix under random orthonormal compression. A matrix manifold is used to model the set of subspace projection matrices. By investigating the differential structure and the condition number of such manifold, we are able to conclude that by $O(s(N-s) \log N)$ random measurements the RIP of subspace projection matrices is guaranteed.

II. MAIN RESULT

In this work, we study the RIP of subspace projection matrices under random orthonormal compression.

Definition 1: The set of projection matrices corresponding to s dimensional subspaces in \mathbb{R}^N is defined as

$$\mathcal{P}_{N,s} := \{ P_X = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} : X \in \mathbb{R}^{N \times s}, \dim(\mathrm{span}(X)) = s \} = \{ P_X = XX^{\mathrm{T}} : X \in \mathrm{Gr}_{N,s} \},$$
(1)

in which span(X) denotes the column space of X, and $\operatorname{Gr}_{N,s}$ is the Grassmann manifold of s dimensional subspaces in \mathbb{R}^N .

Remark 1: Equation (1) is obtained by ortho-normalizing the columns of X while keeping $\operatorname{span}(X)$ fixed. Because different choices of X do not change P_X as long as $\operatorname{span}(X)$ is fixed, we have $X \in \operatorname{Gr}_{N,s}[19]$.

Remark 2: From Definition 1, we know that, for a linear subspace S, its projection matrix is the matrix that a vector $\mathbf{x} \in \mathbb{R}^N$

1070-9908 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

has to multiply when projected onto S. This is the reason that it is called a projection matrix.

 $\mathcal{P}_{N,s}$ is an s(N-s) dimensional submanifold in $\mathbb{R}^{N \times N}$. The following theorem describes the RIP of $\mathcal{P}_{N,s}$ under random orthoprojector.

Theorem 1: For fixed $0 < \varepsilon < 1$ and $\beta > 0$, assume that $N \ge 3$. Let $\mathcal{A} : \mathbb{R}^{N \times N} \to \mathbb{R}^m$ be a random orthoprojector with

$$m \ge \left(\frac{2+\beta}{\varepsilon^2 - \varepsilon^3/3}\right) O\left(s(N-s)\log\frac{N}{\varepsilon}\right).$$
 (2)

If $m < N^2$, then with probability exceeding $1 - e^{-c_1\beta s(N-s)}$, in which c_1 is a universal constant, the following property holds for every pair of $P_X, P_Y \in \mathcal{P}_{N,s}, P_X \neq P_Y$,

$$(1-\varepsilon)\frac{\sqrt{m}}{N} \le \frac{\|\mathcal{A}(P_X - P_Y)\|_2}{\|P_X - P_Y\|_F} \le (1+\varepsilon)\frac{\sqrt{m}}{N}.$$
 (3)

Proof: The proof is postponed to Section IV.

From Theorem 1, the number of measurements $m \ge O(s(N - s) \log N)$ is enough to guarantee the RIP under random orthonormal compression. For low rank matrices without further specific structure, the number of measurements should be no less than $O(sN \log N)$ [4]. One may notice that when $s \ll N$, the improvement from the latter to the former is not much. Although it is true in that case, the result in Theorem 1 does improve the scaling law of the number of measurements on s, and verifies the intuition that, compared with low rank matrices, subspace projection matrices have additional structure which is able to further reduce the number of compressed measurements needed for reconstruction.

Heuristically, the scaling law $m \ge O((N-s)s \log N)$ is reasonable, in that the degree of freedom of a subspace projection matrix is s(N-s). It should be highlighted that although Theorem 1 is for an orthonormal random compression, the conclusion could be naturally extended to random compressions satisfying the concentration property [20].

In the following text, the establishment of Theorem 1 will be demonstrated. The first order differential structure and the condition number of $\mathcal{P}_{N,s}$ are studied in Section III, and then Theorem I is readily proved in Section IV.

III. MANIFOLD OF SUBSPACE PROJECTION MATRIX

Matrix manifold has been a powerful tool to structural matrix data recovery [21]. Typical matrix manifolds, such as the Grassmann manifold, the orthogonal group, and the Stiefel manifold, have been comprehensively studied, and one may read [22], [19], [23] for reference.

In this section, we study the set of subspace projection matrices of s dimensional subspaces in \mathbb{R}^N denoted as $\mathcal{P}_{N,s}$ and defined in (1). Because a subspace has a one to one correspondence with a projection matrix, and such correspondence is continuous, $\mathcal{P}_{N,s}$ is an s(N-s) dimensional manifold which is homeomorphic to $\operatorname{Gr}_{N,s}$.

Preceding the calculation of the differential structure and the condition number, we may first illustrate $\mathcal{P}_{N,s}$ for specific N = 2 and s = 1. From the definition, we know that

$$\mathcal{P}_{2,1} \cong \left\{ \left[x_1^2, \sqrt{2}x_1x_2, x_2^2 \right]^{\mathrm{T}} : \left[x_1, x_2 \right]^{\mathrm{T}} \in \mathrm{Gr}_{2,1} \right\}$$

Because $Gr_{2,1} \cong \mathbb{S}^1$, $\mathcal{P}_{2,1}$ can be expressed as a circle with radius $1/\sqrt{2}$ as shown in Fig. 1.



Fig. 1. $\mathcal{P}_{2,1}$ can be illustrated as a circle with radius $1/\sqrt{2}$.

A. Tangent Space and Normal Space of $\mathcal{P}_{N,s}$

In this part, we study the first order differential structure of $\mathcal{P}_{N,s}$. The reason we need it is that the condition number, which is an important quantity used for the stable embedding of a manifold, will be defined by the normal bundle.

Denote skew(N) as the set of all $N \times N$ skew symmetric matrices, and sym(N) as the set of all $N \times N$ symmetric matrices. In the following lemma, the tangent space and the normal space at every point of $\mathcal{P}_{N,s}$ are unveiled. Remind that by stating $P_X = XX^{\mathrm{T}}$, it is indicated that $X \in \mathrm{Gr}_{N,s}$.

Lemma 1: The tangent space of $\mathcal{P}_{N,s}$ at a point $P_X = XX^{\mathrm{T}}$ is

$$\mathcal{T}_{P_X}\mathcal{P}_{N,s} = \{X_\perp K X^{\mathrm{T}} + X K^{\mathrm{T}} X_\perp^{\mathrm{T}} : K \in \mathbb{R}^{(N-s) \times s}\},$$
(4)

and the normal space is

$$\mathcal{N}_{P_X} \mathcal{P}_{N,s} = \operatorname{skew}(N) \bigcup \{S \in \operatorname{sym}(N) : S = (P_X - P_{X_\perp})S_0, S_0 \in \operatorname{sym}(N)\}.$$
(5)

 X_{\perp} is the matrix such that $[X, X_{\perp}]^{\mathrm{T}}[X, X_{\perp}] = I_N$.

Proof: Denote the sets in (4) and (5) as U and V, respectively, both of which are subspaces.

Suppose that $P_{X(t)} = X(t)X(t)^{\mathrm{T}}$ is a curve on $\mathcal{P}_{N,s}$ with $P_{X(0)} = P_X$. The tangent vector along such curve at point P_X is

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} P_{X(t)} = \frac{\mathrm{d}X(t)}{\mathrm{d}t}\Big|_{t=0} X(0)^{\mathrm{T}} + X(0) \frac{\mathrm{d}X(t)^{\mathrm{T}}}{\mathrm{d}t}\Big|_{t=0}$$
$$= \xi X^{\mathrm{T}} + X\xi^{\mathrm{T}},$$

in which ξ is a vector in $\mathcal{T}_X \operatorname{Gr}_{N,s}$. Because these tangent vectors all belong to the tangent space of $\mathcal{P}_{N,s}$ at P_X , we have

$$\mathcal{T}_{P_X}\mathcal{P}_{N,s} \supset U = \{\xi X^{\mathrm{T}} + X\xi^{\mathrm{T}}: \ \xi \in \mathcal{T}_X \mathrm{Gr}_{N,s}\}$$

To check that $U \perp V$, notice that any symmetric matrix is orthogonal to any skew-symmetric matrix, and that

$$egin{aligned} & ext{tr}((X_{\perp}KX^{ ext{T}}+XK^{ ext{T}}X_{\perp}^{ ext{T}})(XX^{ ext{T}}-X_{\perp}X_{\perp}^{ ext{T}})S_0) \ &= ext{tr}((X_{\perp}KX^{ ext{T}}-XK^{ ext{T}}X_{\perp}^{ ext{T}})S_0) \ &= ext{tr}(X_{\perp}KX^{ ext{T}}S_0^{ ext{T}})- ext{tr}(S_0XK^{ ext{T}}X_{\perp}^{ ext{T}})=0. \end{aligned}$$

The dimension of V is

$$N(N-1)/2 + s(s-1) + (N-s)(N-s-1)/2 + N$$

= N² - Ns + s²,

and the dimension of U is $Ns - s^2$, which is equal to the dimension of $\mathcal{T}_{P_X}\mathcal{P}_{N,s}$. Therefore, we finish the proof that $\mathcal{T}_{P_X}\mathcal{P}_{N,s} = U$ and $\mathcal{N}_{P_X}\mathcal{P}_{N,s} = V$.

B. Condition Number of $\mathcal{P}_{N,s}$

For the purpose of delineating the regularity of a manifold, a notion called the condition number of a manifold, also known as the reach of a manifold [16], is introduced.

Definition 2: [17] Let \mathcal{M} be a compact Riemannian sub-manifold of \mathbb{R}^N . The condition number is defined as $1/\tau$, where τ is the largest number having property that the open normal bundle about \mathcal{M} of radius r is embedded in \mathbb{R}^N for all $r < \tau$.

The condition number $1/\tau$ controls both local properties, such as the curvature of any unit-speed geodesic curve on the manifold, and global properties, such as how close the manifold may curve back upon itself at long geodesic distance [18]. However, from its definition, the condition number of a manifold is not easy to obtain in general. In this section, we focus on the condition number of the manifold $\mathcal{P}_{N,s}$. In the next section, we shall see that the following lemma is a key step to the main result.

Lemma 2: The condition number $1/\tau$ of $\mathcal{P}_{N,s}$ is $\sqrt{2}$.

Proof: The proof is based on the fact that τ is the radius of the largest possible non-self-intersecting tube around the manifold.

The distance between P_X and the set of skew matrices is $||P_X - 0||_F = \sqrt{s}$, in that

$$||P_X||_F = \sqrt{\operatorname{tr}(XX^{\mathrm{T}}XX^{\mathrm{T}})} = \sqrt{\operatorname{tr}(X^{\mathrm{T}}X)} = \sqrt{s}.$$

Thus, if the tube around the manifold intersects itself at a point of skew symmetric matrix, then the radius of the tube is no less than \sqrt{s} .

For any X and Y in $\operatorname{Gr}_{N,s}$, denote $\overline{X} = XX^{\mathrm{T}} - X_{\perp}X_{\perp}^{\mathrm{T}}$ and $\overline{Y} = YY^{\mathrm{T}} - Y_{\perp}Y_{\perp}^{\mathrm{T}}$. If the tube around the manifold intersects itself at a point of symmetric matrix Φ , then $\exists X, Y \in \operatorname{Gr}_{N,s}$ such that

$$\bar{X}S_1 = \bar{Y}S_2 = \Phi \in \operatorname{sym}(N). \tag{6}$$

Equation (6) is equivalent to the condition that Φ and \overline{X} share the same eigenvector matrix \widetilde{X} , and Φ and \overline{Y} share the same eigenvector matrix \widetilde{Y} , which is not necessarily to be the same as \widetilde{X} . One of the eigenspaces of \overline{X} is $\operatorname{span}(X)$ corresponding to eigenvalue 1, and the other one is $\operatorname{span}(X_{\perp})$ corresponding to eigenvalue -1. Thus, \widetilde{X} is composed of basis of $\operatorname{span}(X)$ and basis of $\operatorname{span}(X_{\perp})$. Without loss of generality, we can assume that $\widetilde{X} = [X, X_{\perp}]$ and $\widetilde{Y} = [Y, Y_{\perp}]$. Now we are able to prove that if $\|\Phi - P_X\|_F < 1/\sqrt{2}$ and $\|\Phi - P_Y\|_F < 1/\sqrt{2}$, then (6) can not hold.

First we consider the case where rank(Φ) = s. Suppose that (6) holds, and $\|\Phi - P_X\|_F < 1/\sqrt{2}$ and $\|\Phi - P_Y\|_F < 1/\sqrt{2}$. According to the discussion in the previous paragraph, we have that $\|\Phi - XX^T\|_F = \|\Lambda_1 - \Lambda\|_F$ and $\|\Phi - YY^T\|_F = \|\Lambda_2 - \Lambda\|_F$, in which Λ , Λ_1 , and Λ_2 are diagonal matrices. $\Lambda = \text{diag}[1, \dots, 1, 0, \dots, 0]$, in which the number of 1 is s. Both Λ_1 and Λ_2 have *s* non-zero elements and N - s zeros on the diagonal. Observe that if the non-zeros of Λ_1 is the first *s* elements on its diagonal, then the non-zeros of Λ_2 can not be the first *s* elements on its diagonal. Otherwise, *X* and *Y* span the same subspace. Thus, a contradiction comes from the fact that at least one of $||\Lambda_1 - \Lambda||_F > 1$ and $||\Lambda_2 - \Lambda||_F > 1$ holds.

The second case is rank(Φ) < *s*. If (6) holds, then it is obvious that $\|\Lambda_1 - \Lambda\|_F > 1$ and $\|\Lambda_2 - \Lambda\|_F > 1$, so $\|\Phi - P_X\|_F < 1/\sqrt{2}$ and $\|\Phi - P_Y\|_F < 1/\sqrt{2}$ can not hold.

The third case is rank(Φ) > s. Suppose that the eigenvalues of Φ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. If (6) holds and $\|\Phi - P_X\|_F < 1/\sqrt{2}$, then we must have $\lambda_1 \geq \cdots \geq \lambda_s > 1/2 > \lambda_{s+1} \geq \cdots \geq \lambda_N$, and eigenvalues $\lambda_1, \cdots, \lambda_s$ correspond to the eigenspace span(X). Since span (X) \neq span (Y),

$$\|\Lambda_2 - \Lambda\|_F > \sqrt{(1/2)^2 + (1 - 1/2)^2} = \sqrt{1/2},$$

so $\|\Phi - P_Y\|_F < 1/\sqrt{2}$ can not hold.

These three cases show that $\|\Phi - P_X\|_F < 1/\sqrt{2}$, $\|\Phi - P_Y\|_F < 1/\sqrt{2}$, and (6) can not hold simultaneously. From the discussion in the third case, it is obvious that $\|\Phi - P_X\|_F = 1/\sqrt{2}$, $\|\Phi - P_Y\|_F = 1/\sqrt{2}$, and (6) can hold simultaneously by choosing $\lambda_1 = \cdots = \lambda_{s-1} = 1$, $\lambda_s = \lambda_{s+1} = 1/2$, $\lambda_{s+2} = \cdots = \lambda_N = 0$.

Consequently, for any $P_X, P_Y \in \mathcal{P}_{N,s}, \Phi \in \mathbb{R}^{N \times N}$, if $\|\Phi - P_X\|_F < 1/\sqrt{2}$ and $\|\Phi - P_Y\|_F < 1/\sqrt{2}$, their normal spaces can not intersect at point Φ . If $\|\Phi - P_X\|_F = 1/\sqrt{2}$ and $\|\Phi - P_Y\|_F = 1/\sqrt{2}$, then there exists a Φ at which their normal spaces intersect. Thus, for $\mathcal{P}_{N,s}, \tau = 1/\sqrt{2}$, and the condition number $1/\tau$ is $\sqrt{2}$.

The condition number $1/\tau$ of $\mathcal{P}_{N,s}$ provides the regularity of this manifold, so its RIP is able to be derived.

IV. PROOF OF MAIN RESULT

Proof: Basically, Theorem 1 is proved by applying the condition number of the manifold of the projection matrix $\mathcal{P}_{N,s}$, calculating the covering number of the set of chords of $\mathcal{P}_{N,s}$, and utilizing the Johnson-Lindenstrauss lemma.

The set of chords of a manifold \mathcal{M} is denoted as

$$\mathcal{C}(\mathcal{M}) := \left\{ \frac{X - Y}{\|X - Y\|_F} : \ X, Y \in \mathcal{M}, \ X \neq Y \right\}.$$
(7)

From lemma C.1 in [13], we know that for any $0 < T \le 3\tau/4$, the set $\mathcal{C}(B_T)$ is a $(4\sqrt{T/\tau}, \|\cdot\|_2)$ -cover of $\mathcal{C}(\mathcal{P}_{N,s})$, where B_T is defined as

$$B_T = \bigcup_{P \in \mathcal{N}(\mathcal{P}_{N,s},T)} \{P + \mathcal{T}_P \mathcal{P}_{N,s}(T)\},\tag{8}$$

in which $\mathcal{N}(\mathcal{P}_{N,s},T)$ is the (T,d_g) -cover of $\mathcal{P}_{N,s}$, and $\mathcal{T}_P \mathcal{P}_{N,s}(T) := \{\xi \in \mathcal{T}_P \mathcal{P}_{N,s} : ||\xi||_F \leq T\}$. It is easily shown in Part B of the proof of Theorem III.1 in [13] that the ϵ -cover of $\mathcal{C}(B_T)$ satisfies

$$|\mathcal{N}(\mathcal{C}(B_T),\epsilon)| \leq |\mathcal{N}(\mathcal{P}_{N,s},T)|(1+2/\epsilon)^s + |\mathcal{N}(\mathcal{P}_{N,s},T)|^2(1+2/\epsilon)^{2s+1}.$$
 (9)

Theorem 8 in [24] gives that the (T, d_p) covering number of Grassmann manifold $\operatorname{Gr}_{N,s}$ is $(C_0/T)^{s(N-s)}$ with C_0 being

This paper previously published in IEEE Signal Processing Letters

a universal constant. Remind that the projection distance on Thus, we conclude that if $Gr_{N,s}$ is defined as

$$d_p^2(X,Y) := \frac{1}{2} \|XX^T - YY^T\|_F^2.$$

Thus, the $(\sqrt{2}T, d)$ covering number of $\mathcal{P}_{N,s}$ is also $(C_0/T)^{s(N-s)}$.

For any $Z \in \mathcal{C}(\mathcal{P}_{N,s})$, there exist $\widetilde{Z} \in \mathcal{C}(B_T)$ and $\hat{Z} \in$ $\mathcal{N}(\mathcal{C}(B_T),\epsilon)$ such that $\|(Z-\widetilde{Z})\|_2 \leq 4\sqrt{T/\tau}$ and $\|(\hat{Z}-\tilde{Z})\|_2$ \widetilde{Z}) $\|_2 \leq \epsilon$. We then have

$$\begin{aligned} \|\mathcal{A}(Z)\|_{2} &\leq \|\mathcal{A}(Z - \widetilde{Z})\|_{2} + \|\mathcal{A}(\widetilde{Z} - \hat{Z})\|_{2} + \|\mathcal{A}(\hat{Z})\|_{2} \\ &\leq 4\sqrt{T/\tau} + \epsilon + \|\mathcal{A}(\hat{Z})\|_{2} \end{aligned} \tag{10}$$

and

$$\begin{aligned} \|\mathcal{A}(Z)\|_{2} &\geq \|\mathcal{A}(\hat{Z})\|_{2} - \|\mathcal{A}(Z - \hat{Z})\|_{2} \\ &\geq \|\mathcal{A}(\hat{Z})\|_{2} - (\|\mathcal{A}(Z - \tilde{Z})\|_{2} + \|\mathcal{A}(\tilde{Z} - \hat{Z})\|_{2}) \\ &\geq \|\mathcal{A}(\hat{Z})\|_{2} - 4\sqrt{T/\tau} - \epsilon. \end{aligned}$$
(11)

Equations (10) and (11) together with Lemma 1.1 in [17] (known as the Johnson-Lindenstrauss lemma) give that

$$\sup_{Z \in \mathcal{C}(\mathcal{P}_{N,s})} \|\mathcal{A}(Z)\|_{2} \leq \sup_{\hat{Z} \in \mathcal{N}(\mathcal{C}(B_{T}),\epsilon)} \epsilon + 4\sqrt{T/\tau} + \|\mathcal{A}(\hat{Z})\|_{2}$$
$$\leq \epsilon + 4\sqrt{\frac{T}{\tau}} + (1+\delta)\sqrt{\frac{m}{N^{2}}}, \qquad (12)$$

and

$$\inf_{Z \in \mathcal{C}(\mathcal{P}_{N,s})} \|\mathcal{A}(Z)\|_{2} \geq \inf_{\hat{Z} \in \mathcal{N}(\mathcal{C}(B_{T}),\epsilon)} \|\mathcal{A}(\hat{Z})\|_{2} - (\epsilon + 4\sqrt{T/\tau}) \\
\geq (1-\delta)\sqrt{\frac{m}{N^{2}}} - \epsilon - 4\sqrt{\frac{T}{\tau}}.$$
(13)

Equations (12) and (13) hold simultaneously with probability exceeding $1 - |\mathcal{N}(\mathcal{C}(B_T), \epsilon)|^{-\beta}$ given that

$$m \ge \left(\frac{4+2\beta}{\delta^2/2-\delta^3/3}\right)\log|\mathcal{N}(\mathcal{C}(B_T),\epsilon)|.$$

From Lemma 2, we know that $\tau = 1/\sqrt{2}$. Let $\delta = \varepsilon/2$, $\epsilon =$ $\varepsilon \sqrt{m}/(4N)$, and $T = \tau m \varepsilon^2/(256N^2)$, then

$$\begin{aligned} |\mathcal{N}(\mathcal{C}(B_T),\epsilon)| &\leq \left(\frac{\sqrt{2}C_0}{T}\right)^{s(N-s)} \left(1+\frac{2}{\epsilon}\right)^s \\ &+ \left(\frac{\sqrt{2}C_0}{T}\right)^{2s(N-s)} \left(1+\frac{2}{\epsilon}\right)^{2s+1} \\ &\leq 2 \left(\frac{512N^2C_0}{m\epsilon^2}\right)^{2s(N-s)} \left(1+8\sqrt{\frac{N^2}{m\epsilon^2}}\right)^{2s+1} \\ &\leq 2 \left(\frac{512N^2C_0}{\epsilon^2}\right)^{2s(N-s)} \left(1+8\frac{N}{\epsilon}\right)^{2s+1}. \end{aligned}$$

$$\begin{split} m &\geq \left(\frac{4+2\beta}{\varepsilon^2/8 - \varepsilon^3/24}\right) \left(\log 2 + 2s(N-s)\log\left(\frac{512N^2C_0}{\varepsilon^2}\right) \\ &+ (2s+1)\log\left(1 + \frac{8N}{\varepsilon}\right)\right) \\ &\sim \left(\frac{2+\beta}{\varepsilon^2 - \varepsilon^3/3}\right) O\left(s(N-s)\log\left(\frac{N}{\varepsilon}\right)\right) \end{split}$$

then $\forall Z \in \mathcal{C}(\mathcal{P}_{N,s}), (1 - \varepsilon)\sqrt{m}/N \leq ||\mathcal{A}(Z)||_2$ $(1 + \varepsilon)\sqrt{m}/N$ holds with probability exceeding 1 $|\mathcal{N}(\mathcal{C}(B_T),\epsilon)|^{-\beta}.$

In order to control the probability above, we need a lower bound on $|\mathcal{N}(\mathcal{C}(B_T), \epsilon)|$. Notice that

$$\begin{split} |\mathcal{N}(\mathcal{C}(B_T),\epsilon)| &\geq |\mathcal{N}(\mathcal{P}_{N,s},T)| \left(1 + \frac{2}{\epsilon}\right)^s \\ &\geq |\mathcal{N}(\mathcal{P}_{N,s},T)| = |\mathcal{N}\left(\operatorname{Gr}_{N,s},\frac{T}{\sqrt{2}}\right)| \\ &\geq \left(\frac{\sqrt{2}c_0}{T}\right)^{s(N-s)}, \end{split}$$

in which the last inequality holds when $T \leq \sqrt{2}\pi/4$ according to Theorem 8 in [24], and c_0 is a universal constant. Because T $= \tau m \varepsilon^2 / (256 N^2) < \tau / 256 = 1 / (256 \sqrt{2})$, we have

$$|\mathcal{N}(\mathcal{C}(B_T),\epsilon)| \ge (512c_0)^{s(N-s)} = \exp(c_1 s(N-s)).$$

Thus, the probability exceeding

$$1 - |\mathcal{N}(\mathcal{C}(B_T), \epsilon)|^{-\beta} \ge 1 - \exp(-c_1\beta s(N-s)).$$

V. CONCLUSION

Subspace projection matrices are low rank matrices with additional structure that allows for further reduction of its intrinsic dimension. In this work, the restricted isometry property of subspace projection matrix under random orthonormal compression is studied.

The set of s dimensional subspace projection matrices $\mathcal{P}_{N,s}$ is modeled as an s(N-s) dimensional submanifold in $\mathbb{R}^{N \times N}$ so the main concern is transformed to the problem of the stable embedding of $\mathcal{P}_{N,s}$ into $\mathbb{R}^{N \times N}$. One of the key points is the calculation of the conditional number $1/\tau$ of $\mathcal{P}_{N,s}$. Once τ is obtained, the RIP is able to be established by applying covering sets of the set of chords of $\mathcal{P}_{N,s}$ and utilizing the JL lemma. In order to calculate the condition number $1/\tau$, the tangent space and the normal space at every point of $\mathcal{P}_{N,s}$ are investigated. The result is that by $O(s(N-s)\log N)$ measurements the RIP of subspace projection matrix is guaranteed.

This work is not exhausted. The condition number $1/\tau$ of $\mathcal{P}_{N,s}$ provides the regularity of this manifold, so its RIP under other random compressions, such as i.i.d. Gaussian compression, could also be established using such condition number. Furthermore, theoretical analysis on algorithms for compressed subspace projection matrix recovery could be built upon this work.

This paper previously published in IEEE Signal Processing Letters

REFERENCES

- E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [2] M. Fazel, "Matrix Rank Minimization with Applications," PhD thesis, Stanford Univ., Stanford, CA, USA, 2002.
- [3] R. Baraniuk, M. Davenport, R. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
- [4] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [5] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.
- [6] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Math.*, vol. 346, no. 9, pp. 589–592, 2008.
- [7] K. C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.
- [8] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," in *IEEE Conf. Computer Vision and Pattern Recognition, 2008 CVPR*, Oct. 2010, pp. 1–8.
- [9] J. Ho, K. C. Lee, M. H. Yang, and D. Kriegman, "Visual tracking using learned linear subspaces," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2004 CVPR, Jun. 2004, vol. 1, pp. I–782–I–789.
- [10] L. Balzano, R. Nowak, and B. Recht, "Online identification, and tracking of subspaces from highly incomplete information," in 48th Annu. Allerton Conf. Communication, Control, and Computing, 2010, Sep. 2010, pp. 704–711.
- [11] Y. Chi, Y. C. Eldar, and R. Calderbank, "Petrels: Subspace estimation, and tracking from partial observations," in *IEEE Int. Conf. Acoustics*, *Speech and Signal Processing (ICASSP)*, 2012, Mar. 2012.

- [12] X. Mao and Y. Gu, "Compressed subspace clustering: A case study," in *IEEE Global Conf. Signal and Information Processing*, Dec. 2014.
- [13] H. L. Yap, M. B. Wakin, and C. J. Rozell, "Stable manifold embeddings with structured random matrices," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 4, pp. 720–730, Aug. 2013.
- [14] A. Eftekhari and M. B. Wakin, "New analysis of manifold embeddings and signal recovery from compressive measurements," *CoRR*, 2013, abs/1306.4748.
- [15] M. A. Iwen and M. Maggioni, "Approximation of points on low-dimensional manifolds via random linear projections," *CoRR*, 2012, abs/ 1204.3337.
- [16] K. L. Clarkson, "Tighter bounds for random projections of manifolds," in Symp. Computational Geometry, 2008, pp. 39–48, ACM.
- [17] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Computat. Math.*, vol. 9, no. 1, pp. 51–77, 2009.
- [18] P. Niyogi, S. Smale, and S. Weinberger, "Finding the homology of submanifolds with high confidence from random samples," *Discr. Computat. Geom.*, vol. 39, no. 1 3, pp. 419–441, 2008.
- [19] P. A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds. Princeton, NJ, USA: Princeton Univ. Press, 2008.
- [20] D. Achlioptas, "Database-friendly random projections: Johnson-lindenstrauss with binary coins," *J. Comput. Syst. Sci*, vol. 66, no. 4, pp. 671–687, 2003.
- [21] M. Tan, I. W. Tsang, L. Wang, B. Vandereycken, and S. J. Pan, "Riemannian pursuit for big matrix recovery," in *Proc. 31st Int. Conf. Machine Learning (ICML) volume of JMLR Workshop and Conf. Proc.*, 2014, pp. 1539–1547.
- [22] J. M. Lee, Introduction to Smooth Manifolds. Berlin, Germany: Springer, 2003.
- [23] P. A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of grassmann manifolds with a view on algorithmic computation," *Acta Appl. Math.*, vol. 80, no. 2, pp. 199–220, 2004.
- [24] S. J. Szarek, "Metric entropy of homogeneous spaces," Quant. Probabil., Banach Center Publ. 43, Polish Acad. Sci, pp. 395–410, 1998.