Subspace Projection Matrix Recovery from Incomplete Information

Xinyue Shen and Yuantao Gu

Abstract

Structural signal retrieval from highly incomplete information is a concerning problem in the field of signal and information processing. In this work, we study the recovery of subspace projection matrix from random compressed measurements, i.e., matrix sensing, and random down-samplings, i.e., matrix completion, by formulating an optimization problem on the Grassmann manifold. For the sensing problem, we derive a bound on the number of Gaussian measurements $O(s(N-s))$, so that a restricted isometry property can hold with high probability. Such RIP condition can guarantee the unique recovery in the noiseless scenario, and the robust recovery in the noise case. As for the matrix completion problem, we obtain a bound on the sampling density of the Bernoulli model $O\left(s^{3/2}(N-s)\log^3 N/\left(N^2\right)\right)$ for subspace projection matrices. A gradient descent algorithm on the Grassmann manifold is proposed to solve the mentioned optimization problem, and the convergence behavior of such non-convex optimization algorithm is theoretical analyzed for the sensing and the completion problem respectively. The algorithm is numerically tested in both the sensing and the completion problems under both noiseless and noise scenarios. Theoretical results are verified, and the algorithm is compared with some other low rank matrix completion algorithms to demonstrate its good performance.

Index Terms

Subspace projection matrix, matrix sensing, matrix completion, gradient descent on the Grassmann manifold

X. Shen and Y. Gu are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: shenxy13@mails.tsinghua.edu.cn. gyt@tsinghua.edu.cn.

Subspace Projection Matrix Recovery from Incomplete Information

I. Introduction

Structural signal retrieval from highly incomplete information has always been a concerning problem in the field of signal and information processing. Sparse signal and signal on low dimensional manifold have been important structural models in both real-world applications [1]–[4] and theories of compressive sensing [5]–[9] and compressive estimation [10], [11]. At the same time, low rank matrix as another commonly encountered structural data [12]–[14] has attracted numerous studies in matrix completion [15], [16] and matrix sensing [17], [18]. These structures invoke low intrinsic dimension which leaves room for non-adaptive dimensionality reduction and compressive recovery [19], so theories and methods can be established to achieve exact or robust reconstructions from incomplete information.

Subspace projection matrix is also a kind of structural data. For an s dimensional linear subspace $\mathcal{S}$ in an Euclidean space $\mathbb{R}^N$, the subspace projection matrix is the matrix that a vector has to multiply when projected onto $\mathcal{S}$. The rank of such subspace projection matrix is s, so when $s \ll N$ it is a low rank matrix. However, it has more specific structure. It is symmetric and semi-definite, and its only non-zero eigenvalue is 1. Such additional structure invokes intrinsic dimension $s(N-s)$, which is lower than that of a general fixed rank matrix. In work [20], the restricted isometry property of subspace projection matrices under random ortho-projectors has been established. Such result further indicates that theoretical improvements and more efficient algorithms can be expected for subspace projection matrix sensing and completion.

Subspace projection matrix has natural connections to manifolds. Firstly, the projection matrices of subspaces of a fixed dimension s in $\mathbb{R}^N$ form a manifold [20]. There have existed brilliant and solid works in compressive recovery of signals on low dimensional manifold [9], [21]–[24]. These works mainly focus on stable manifold embeddings, and extend classic compressive sensing by generalizing the low dimension model from sparse signal to signal on low dimensional manifold. Especially, a key quantity called the condition number of a manifold is utilized to study submanifold extrinsically and unveil important Riemannian geometry properties. By controlling the regularity of a manifold via such quantity, theorems on the RIP condition are established accordingly.

Secondly, the set of all s dimensional subspaces in $\mathbb{R}^N$ form the Grassmann manifold. There have existed works in retrieving structural data by solving optimization problems on manifolds. The work [25] studies complete dictionary recovery, and a Riemannian trust region algorithm is proposed to solve a non-convex optimization problem with a spherical constraint. In [26], a conjugate gradient descent method on the Grassmann manifold is proposed to estimate the projection matrix and the intercept vector of an affine subspace from sample points contaminated by noise. The work [27] solves the problem of online subspace identification and tracking from incomplete information by incremental gradient descent on the Grassmann manifold. The problem of low-rank matrix completion is also studied in the work [28] by formulating a minimization problem on the Riemannian manifold of matrices with fixed rank.

A. Motivation

In this work, we study the recovery of subspace projection matrix from incomplete information, which includes random compressed measurements and random down-samplings of the matrix. The former is called the subspace projection matrix sensing problem, and the latter is subspace projection matrix completion problem.

This work is mainly motivated from two aspects. The first one is to explore whether the specific structure of subspace projection matrix can bring benefits to dimensionality reduction and compressive recovery. It is well known that for low rank matrix of size $N \times N$ and rank s, the number of measurements should be $O(sN)$ to guarantee a restricted isometry property under random Gaussian compression [17], and the density of the Bernoulli random sampling should be $O\left(\frac{s^2 \log^3 N}{N}\right)$ to guarantee matrix completion for $\log N$-incoherent matrices [29]. In this work, we would like to explore theoretical guarantees for a special kind of fixed rank matrix, the subspace projection matrix. Due to the reason that the dimension of freedom of a subspace projection matrix is lower than that of a general fixed rank matrix, theoretical improvements are expected.

The second motivation is from the view of application. Subspace projection matrix has a one to one correspondence with subspace, so its recovery from incomplete information can be viewed as subspace estimation. Subspace estimation has been an interesting problem in face recognition [30], motion segmentation [31], and visual tracking [32], and more recently subspace estimation from highly incomplete information has appeared as an attractive research topic [33]–[35]. Subspace projection matrix itself is an important kind of structural matrix. For example, a sequence of signal passing a multi-band-pass filter is equivalent to a subspace projection matrix multiplying the signal vector, so a bandpass filter can be implemented by the corresponding subspace projection matrix. Given such a filter, it would be of interests to estimate its pass bands from highly incomplete information such as random partial entries.

B. Problem formulation

For a matrix $\mathbf{X} \in \mathbb{R}^{N\times s}$ of rank s, the projection matrix corresponding to its column space $\mathrm{span}(\mathbf{X})$, which is an s dimensional subspace in $\mathbb{R}^N$, is defined as

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T} = \bar{\mathbf{X}}\bar{\mathbf{X}}^\mathsf{T},\ \bar{\mathbf{X}} \in \mathrm{Gr}_{N,s}, \tag{1}$$

in which $\mathrm{Gr}_{N,s}$ is the Grassmann manifold of s dimensional subspaces in $\mathbb{R}^N$, and $\bar{\mathbf{X}}$ is obtained by ortho-normalizing the columns of $\mathbf{X}$.

In this work, $\mathcal{A} : \mathbb{R}^{N\times N} \to \mathbb{R}^M$ is a random mapping

$$[\mathcal{A}(\mathbf{P}_{\mathbf{X}})]_j := \langle \mathbf{P}_{\mathbf{X}}, \mathbf{A}_j\rangle,\ j = 1, \cdots, M, \tag{2}$$

in which $\{\mathbf{A}_j \in \mathbb{R}^{N\times N}\}_{j=1}^M$ are independent random matrices. In the case of matrix sensing, $M < N^2$, and each $\mathbf{A}_j$ is a random Gaussian matrix. In the case of matrix completion, $M = N^2$, all but the j-th elements[1] of $\mathbf{A}_j$ are zero, and the j-th element follows the Bernoulli model. The measurement vector $\mathbf{y} \in \mathbb{R}^M$ is

$$\mathbf{y} = \mathcal{A}(\mathbf{P}_{\mathbf{X}_0}) + \mathbf{w}, \tag{3}$$

in which $\mathbf{w} \in \mathbb{R}^M$ is an additive noise. The aim is to recover the subspace $\mathrm{span}(\mathbf{X}_0)$ from $\mathbf{y}$ by the following minimization problem on the Grassmann manifold

$$\min_{\mathbf{X}\in\mathrm{Gr}_{N,s}} \|\mathcal{A}(\mathbf{P}_{\mathbf{X}}) - \mathbf{y}\|_2^2. \tag{4}$$

C. Contribution and paper organization

To begin with, an optimization problem on the Grassmann manifold is formulated as (4). In the next section, some useful preliminaries will be reviewed to make this paper self-contained. In section III, we will study the subspace projection matrix sensing problem, and the mapping $\mathcal{A}$ will be the random Gaussian compression. We derive that with the number of Gaussian measurements M of the magnitude of $s(N - s)$, a restricted isometry property can hold with high probability. Such RIP condition can

[1]For an $N \times N$ matrix $\mathbf{P}$, its j-th element denoted as $[\mathbf{P}]_j$ is $[\mathbf{P}]_{a,b}$ with $j = N(b - 1) + a$.

guarantee the unique recovery of problem (4) in the noiseless scenario as well as the robust recovery for the noise case. In section IV, we will deal with the subspace projection matrix completion problem by assuming the mapping $\mathcal{A}$ to obey the random Bernoulli model, and obtain a theoretical bound on the sampling density, so that an RIP-like condition can hold with high probability. In section V, we propose to solve problem (4) via a gradient descent algorithm on the Grassmann manifold. More importantly, the convergence behavior of such non-convex optimization algorithm is theoretical analyzed for the sensing and the completion problem respectively. Numerical experiments will be demonstrated in section VI. Simulations verify that the number of measurements M for the sensing problem should be $1.6s(N-s)$, and the density of the Bernoulli sampling in the completion problem should be $1.8s(N-s)/N^2$ in the noiseless scenario when $N=20$. The robustness of the algorithm is also tested. In the completion problem, the proposed algorithm is compared with some other low rank matrix completion algorithms.

II. Preliminary

In order to make this paper self-contained, we have a brief review on the Grassmann manifold, the condition number of the manifold of subspace projection matrices, and the line-search algorithm on manifolds.

A. Grassmann manifold

The Grassmann manifold of s dimensional subspaces in $\mathbb{R}^N$, denoted as $\mathrm{Gr}_{N,s}$, can be defined as the following quotient manifold

$$\mathrm{Gr}_{N,s} := \frac{\mathcal{O}(N)}{\mathcal{O}(s)\times\mathcal{O}(N-s)}, \tag{5}$$

in which $\mathcal{O}(N)$ it the $N\times N$ orthogonal matrix group. $\mathrm{Gr}_{N,s}$ is a compact differentiable manifold of dimension $d=s(N-s)$. Any s dimensional subspace $\mathcal{X}\subset\mathbb{R}^N$ is a point on $\mathrm{Gr}_{N,s}$, and we can represent it by a matrix $\mathbf{X}\in\mathbb{R}^{N\times s}$ such that the columns of $\mathbf{X}$ span $\mathcal{X}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}=\mathbf{I}_s$.

The tangent space of $\mathrm{Gr}_{N,s}$ at a point $\mathbf{X}$ is

$$\mathcal{T}_{\mathbf{X}}\mathrm{Gr}_{N,s} = \{\boldsymbol{\xi}\in\mathbb{R}^{N\times s} : \mathbf{X}^\mathsf{T}\boldsymbol{\xi}=\mathbf{0}\}.$$

For any tangent vectors $\boldsymbol{\eta},\boldsymbol{\xi}\in\mathcal{T}_{\mathbf{X}}\mathrm{Gr}_{N,s}$, the metric on $\mathrm{Gr}_{N,s}$ induced by the Euclidean metric in $\mathbb{R}^N$ is

$$\langle\boldsymbol{\eta},\boldsymbol{\xi}\rangle_{\mathbf{X}} := \mathrm{trace}\left((\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{\xi}\right), \tag{6}$$

and the geodesic distance induced by such metric is

$$d_g^2(\mathbf{X}, \mathbf{Y}) := \sum_{i=1}^{s} \theta_i^2,$$

where θ_i are the principal angles between these two subspaces. Besides the geodesic distance, the projection distance is defined by the subspace projection matrix

$$d_p^2(\mathbf{X}, \mathbf{Y}) := \frac{1}{2}\|\mathbf{P_X} - \mathbf{P_Y}\|_F^2 = \sum_{i=1}^{s} \sin^2 \theta_i.$$

B. Manifold of subspace projection matrices and its condition number

Before looking into the manifold of subspace projection matrices, let us first introduce the condition number of a manifold, also known as the reach of a manifold [24], which 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 [36].

Definition 1. *[9] 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 $r < \tau$.*

For more detailed knowledge about the condition number, please refer to [9]. Next, let us define the set of subspace projection matrices of rank s

$$\mathcal{P}_{N,s} := \{\mathbf{X}\mathbf{X}^\mathsf{T}, \mathbf{X} \in \mathrm{Gr}_{N,s}\}.$$

In one of our recent works [20], we have achieved the following result on $\mathcal{P}_{N,s}$.

Lemma 1. *[20] $\mathcal{P}_{N,s}$ is an $s(N-s)$ dimensional manifold. Its tangent space at a point $\mathbf{P_X} = \mathbf{X}\mathbf{X}^\mathsf{T}$ is*

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

in which $\mathbf{X}_\perp$ is a matrix such that $[\mathbf{X}, \mathbf{X}_\perp]^\mathsf{T}[\mathbf{X}, \mathbf{X}_\perp] = \mathbf{I}_N$, and condition number τ of $\mathcal{P}_{N,s}$ is $\frac{1}{\sqrt{2}}$.

The reason that we need the condition number of $\mathcal{P}_{N,s}$ is that it controls the manifold regularity [9], so we are able to construct an ϵ covering net of the set of chords of $\mathcal{P}_{N,s}$. Our theoretical conclusions will be based on such result.

C. General framework of line-search algorithm on manifold

The line-search algorithm on Riemannian manifold $\mathcal{M}$ is generalized from the iterative line-search method in an Euclidean space. In the k^{th} iteration, a tangent vector $\boldsymbol{\eta}_{k+1}$ to $\mathcal{M}$ at the previous iterate $\mathbf{X}_k$ is selected, and a search along a curve $\gamma(t) \subset \mathcal{M}$ such that $\gamma'(0) = \boldsymbol{\eta}_{k+1}$ is performed.

In order to find the descent direction of the cost function f on the tangent space at a point $\mathbf{X}$, we need the notion of the gradient of f with respect to $\mathcal{M}$.

Definition 2. *[37] Given a smooth scalar field f on a Riemannian manifold $\mathcal{M}$, the gradient of f at a point $\mathbf{X} \in \mathcal{M}$ denoted as $\mathrm{grad}f(\mathbf{X})$ is defined as the unique element in $\mathcal{T}_{\mathbf{X}}\mathcal{M}$ such that*

$$\langle \mathrm{grad}f(\mathbf{X}), \boldsymbol{\xi} \rangle_{\mathbf{X}} = \nabla f(\mathbf{X})[\boldsymbol{\xi}], \ \ \forall \boldsymbol{\xi} \in \mathcal{T}_{\mathbf{X}}\mathcal{M},$$

in which the linear mapping $\nabla f(\mathbf{X})$ is the differential of f at $\mathbf{X}$.

Any direction $\boldsymbol{\xi} \in \mathcal{T}_{\mathbf{X}}\mathcal{M}$ with $\langle \boldsymbol{\xi}, \mathrm{grad}f(\mathbf{X}) \rangle_{\mathbf{X}} < 0$ is a gradient related descent direction. Next, a curve along a gradient related descent direction is defined by the retraction map, which retracts a vector in the tangent space back onto the manifold.

Definition 3. *[37] A retraction on a manifold $\mathcal{M}$ is a smooth mapping R from the tangent bundle $\mathcal{T}\mathcal{M}$ onto $\mathcal{M}$ with the following properties. Let $R_{\mathbf{X}}$ denote the restriction of R to $\mathcal{T}_{\mathbf{X}}\mathcal{M}$.*

1) $R_{\mathbf{X}}(0) = \mathbf{X}$,
2) $\nabla R_{\mathbf{X}}(0) = \mathrm{id}_{\mathcal{T}_{\mathbf{X}}\mathcal{M}}$,

in which $\mathrm{id}_{\mathcal{T}_{\mathbf{X}}\mathcal{M}}$ is the identity map on $\mathcal{T}_{\mathbf{X}}\mathcal{M}$.

One way to define the retraction map on $\mathrm{Gr}_{N,s}$ is simply by the Q factor of the QR decomposition [37]

$$R_{\mathbf{X}}(\boldsymbol{\xi}) = \mathrm{qr}(\mathbf{X} + \boldsymbol{\xi}).$$

Given that a retraction map is well defined, and a sequence of gradient related directions can be numerically computed, such general line-search framework on non-linear manifold can be an efficient solver of optimization problem on manifold.

III. Subspace Projection Matrix Sensing

In this section, we will study the subspace projection matrix sensing problem, and each $\mathbf{A}_j$ in (2) is assumed to have independent Gaussian entries

$$[\mathbf{A}_j]_{k,l} \sim \mathcal{N}(0, 1/M), \quad 1 \le j \le M, 1 \le k, l \le N. \tag{7}$$

The aim is to recover $\mathrm{span}(\mathbf{X}_0)$ from $\mathbf{Y}$ by solving the minimization problem (4).

A. Unique recovery in noiseless cases

For the recovery of the subspace projection matrix from random Gaussian compression, we tend to find the theoretical condition under which the unique solution to problem (4) is precisely $\mathbf{X}_0$ when there is no noise. Since $\mathbf{X}_0$ is already a global minimum, we only need to prove the uniqueness of the global minimum of problem (4). Such uniqueness can be established upon the following theorem on the restricted isometry property of subspace projection matrices under random Gaussian compression.

Theorem 1. *Fix $0 < \epsilon < 1/3$, $0 < \rho < 1$, and $N > s > 0$. Let $\mathcal{A} : \mathbb{R}^{N\times N} \to \mathbb{R}^M$ be a random compression operator with independent $\{\mathbf{A}_j\}_{j=1}^M$ following the distribution in (7).*

$$M \ge 18\epsilon^{-2} \cdot \max\left\{2s(N-s)\log\left(\frac{c_0}{\epsilon^2}\right), \log\left(\frac{8}{\rho}\right)\right\}, \tag{8}$$

in which c_0 is a universal constant. If $M < N^2$, then with probability exceeding $1-\rho$, the following property holds for every pair of $\mathbf{P_X} \ne \mathbf{P_Y} \in \mathcal{P}_{N,s}$,

$$1 - \epsilon \le \frac{\|\mathcal{A}(\mathbf{P_X} - \mathbf{P_Y})\|_2}{\|\mathbf{P_X} - \mathbf{P_Y}\|_F} \le 1 + \epsilon. \tag{9}$$

Proof. The proof is postponed to Appendix A. □

Remark 1. *If*

$$\rho \ge 8\left(\frac{\epsilon^2}{c_0}\right)^{2s(N-s)}, \tag{10}$$

then $2s(N-s)\log\left(c_0/\epsilon^2\right) \ge \log(8/\rho)$, and (8) becomes

$$M \ge \frac{36}{\epsilon^2}\log\left(\frac{c_0}{\epsilon^2}\right)s(N-s),$$

which makes the scaling law more clear. To satisfy (10), if $N^2 > \max\left(36\mathrm{e}s(N-s)/c_0, 324\log(9c_0)s(N-s)\right)$, we can choose $\epsilon^2 = \min(c_0/\mathrm{e}, 1/9)$, and $\rho = 8\mathrm{e}^{-2s(N-s)}$, so the probability that (9) holds exceeds $1 - 8\mathrm{e}^{-2s(N-s)}$.

Given that (9) holds, two different subspace projection matrices will not be mapped into the same image, so the unique global minimum of problem (4) is $\mathbf{X}_0$ when there is no noise. The bound in (8) is tight in the sense that the number of measurements M required to guarantee the restricted isometry property with high probability is of the magnitude of $s(N-s)$, which is exactly the degrees of freedom of subspace projection matrices of rank s. In work [20], we have proved that if the random compression is a random orthogonal operator, then $M \geq O(s(N-s)\log N)$ is required. Here, we show that with Gaussian random compression, the $\log N$ term can be further eliminated.

Classic result in low rank matrix sensing states that the number of measurements M should be $O(sN)$ [17]. Theorem 1 improves the scaling law of M in s for subspace projection matrices, and when s grows, such improvement becomes significant. More importantly, Theorem 1 clearly verifies the intuition that compared with fixed rank matrices the lower intrinsic dimension of subspace projection matrices is able to further reduce the number of compressed measurements needed for reconstruction.

B. Robustness

It is well known that when recovering a sparse signal or a low rank matrix from a convex optimization problem, the bound on the reconstruction error in the noise scenario has to be established upon complicated proofs [17], [38]. However, by the proposed non-convex optimization problem, the robustness is a rather direct result, given that the restricted isometry property holds.

Proposition 1. *Suppose that* $\mathbf{y} = \mathcal{A}(\mathbf{P}_{\mathbf{X}_0}) + \mathbf{w}$*, and equation* (9) *holds. Then the solution* $\mathbf{P}_{\hat{\mathbf{X}}}$ *satisfies*

$$\|\mathbf{P}_{\hat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}_0}\|_F \leq 2\frac{\|\mathbf{w}\|_2}{1-\epsilon}, \tag{11}$$

in which ϵ *is defined in Theorem 1.*

Proof. The proof comes directly from the observation that

$$\begin{aligned}
\|\mathbf{P}_{\hat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}_0}\|_F &\leq \frac{\|\mathcal{A}(\mathbf{P}_{\hat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}_0})\|_2}{1-\epsilon} \\
&= \frac{\|\mathcal{A}(\mathbf{P}_{\hat{\mathbf{X}}}) - \mathbf{y} + \mathbf{w}\|_2}{1-\epsilon} \\
&\leq \frac{\|\mathcal{A}(\mathbf{P}_{\hat{\mathbf{X}}}) - \mathbf{y}\|_2 + \|\mathbf{w}\|_2}{1-\epsilon} \\
&\leq \frac{\|\mathcal{A}(\mathbf{P}_{\mathbf{X}_0}) - \mathbf{y}\|_2 + \|\mathbf{w}\|_2}{1-\epsilon} \\
&= 2\frac{\|\mathbf{w}\|_2}{1-\epsilon}.
\end{aligned}$$

The third inequality is from the fact that $\mathbf{P}_{\hat{\mathbf{X}}}$ is the solution to problem (4). □

The robustness stated in Proposition 1 comes from the constraint that the solution is on the manifold $\mathrm{Gr}_{N,s}$. This result also indicates that the solution to the non-convex problem (4) is highly dependent on the RIP, and when using the Gaussian random compression, with a proper choice of M, the RIP can hold with high probability, so the quality of the solution can be guaranteed.

IV. Subspace Projection Matrix Completion

Given that the mapping $\mathcal{A}$ is a random down-sampling operator, the problem becomes matrix completion from partial entries. In this section, we will study the subspace projection matrix completion problem, and the independent $\{\mathbf{A}_j\}_{j=1}^{N^2}$ are assumed to follow the Bernoulli random down-sampling model with $0 < \rho < 1$

$$\begin{aligned} &\mathbb{P}\{[\mathbf{A}_j]_{k,l} = 0\} = 1,\ k + (N-1)l \neq j \\ &\mathbb{P}\{[\mathbf{A}_j]_{k,l} = 1\} = \rho,\ k + (N-1)l = j \\ &\mathbb{P}\{[\mathbf{A}_j]_{k,l} = 0\} = 1 - \rho,\ k + (N-1)l = j. \end{aligned} \tag{12}$$

Note that the average number of non-zero samples in the above Bernoulli model is ρN^2.

First we introduce the definition of α-regular matrix.

Definition 4. *[29] If a matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$ satisfies that*

$$\|\mathbf{P}\|_\infty := \max_{1\le i,j\le N} |\mathbf{P}_{ij}| \le \frac{\alpha\|\mathbf{P}\|_F}{N},$$

then $\mathbf{P}$ is α-regular.

We know that $1 \le \alpha \le N$, and if a matrix satisfies $\alpha \ll N$, then its F-norm is not contributed by only a few entries, i.e., it is not sparse. A matrix with small α is preferred in a matrix completion problem, in that if α is large, then only a few entries have significant values, so it would be highly likely that the down-sampling procedure only collects small or zero entries. Hence, such condition is crucial in the theory of matrix completion [29].

In the following, we establish an RIP-like bound on the sampling probability ρ for the Bernoulli model in (12) for subspace projection matrices.

Theorem 2. *Fix $\beta > 1$, $\alpha \geq 1$, $0 < \varepsilon < 1$, and $N > s > 0$. For $\mathcal{A}$ defined according to the Bernoulli model in* (12) *with density*

$$\rho \geq \frac{12\alpha\beta}{N^2\varepsilon^2}\Bigg[\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right) + 2C_1 s(N-s)\log\left(\frac{N^2\varepsilon}{12\alpha\beta}\right)\Bigg] \tag{13}$$

and N large enough so that $\rho < 1$, in which C_1 is a constant, then with probability at least $1 - 2\mathrm{e}^{-\beta}$, the following inequalities hold for every pair of $\mathbf{P_X} \neq \mathbf{P_Y} \in \mathcal{P}_{N,s}$ such that $\mathbf{P_X} - \mathbf{P_Y}$ is α-regular

$$\sqrt{1-\frac{\varepsilon}{2}} - \sqrt{\frac{\varepsilon}{2}} \leq \frac{\|\mathcal{A}(\mathbf{P_X}-\mathbf{P_Y})\|_2}{\sqrt{\rho}\|\mathbf{P_X}-\mathbf{P_Y}\|_F} \leq \sqrt{1+\frac{\varepsilon}{2}} + \sqrt{\frac{\varepsilon}{2}}. \tag{14}$$

Proof. The proof is postponed to Appendix B. □

Remark 2. *Because $0 < \varepsilon < 1$, the left hand side of* (14) *is always positive.*

Remark 3. *The parameters β and ε are arbitrary, and not necessarily dependent on N or s. Yet the parameter α is related to N and s. According to Lemma 2.4 in [29], if a matrix of rank s is $\alpha/\sqrt{s}$-incoherent, then it is α-regular. From [39], if the eigen-space is generated from the uniform distribution or spanned by vectors with bounded entries, then its coherence parameter is no larger than $\log N$ with high probability. Hence, α is not larger than $\sqrt{2s}\log N$ with high probability.*

The bound in (13) is of the magnitude of $O\left(\alpha\beta s(N-s)\log(N\sqrt{\varepsilon})/\left(N^2\varepsilon^2\right)\right)$. If we plug in $\alpha = \sqrt{2s}\log N$ and $\beta = \log N$, then it becomes $O\left(s^{3/2}(N-s)(\log N)^2\log(N\sqrt{\varepsilon})/\left(N^2\varepsilon^2\right)\right)$. Classic result in low rank matrix completion (Theorem 2.2 in [29]) states that, for Bernoulli sampling model, the density should be no less than $O\left(s^2(\log N)^3/\left(\delta^2 N\right)\right)$ to guarantee an RIP-like inequality for $\log N$-incoherent matrices, in which δ is the constant in the RIP-like inequality. The bound on the density in Theorem 2 reduces from s^2/N to $s^{3/2}(N-s)/N^2$, due to that the intrinsic dimension reduces to $s(N-s)$.

V. Gradient Descent Method on Grassmann Manifold

We propose to use a manifold gradient descent method to solve problem (4). The algorithm called Grassmann manifold gradient descent line search method (GGDLS) is described in Table I. [2]

[2] We have previously proposed GGDLS in a conference paper [40] to solve the subspace projection matrix completion with the uniformly random down-sampling model, but we did not give theoretical analysis. In this work, we re-mention this algorithm, and use it to solve both the sensing problem and the completion problem with the Bernoulli down-sampling model. More importantly, we include theoretical analyses in this work.

In the first step, the Euclidean gradient $\nabla f_{\mathbf{X}_k}$ is an $N \times s$ matrix, and its entries are computed as

$$[\nabla f(\mathbf{X})]_{a,b} = (\mathcal{A}(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_0}))^\mathsf{T} \cdot \widetilde{\mathbf{A}_a} \cdot \mathbf{x}_b,$$
$$\forall a = 1, 2, \cdots, N, b = 1, 2, \cdots, s,$$

in which $\widetilde{\mathbf{A}_a}$ is an $M \times N$ matrix with entries $[\widetilde{\mathbf{A}_a}]_{j,k} = [\mathbf{A}_j]_{k,a} + [\mathbf{A}_j]_{a,k}$, and $\mathbf{x}_b$ is the b-column of $\mathbf{X}$.

In the second step, $\boldsymbol{\eta}_k$ is the projection of $\nabla f_{\mathbf{X}_k}$ onto the tangent space of $\mathrm{Gr}_{N,s}$ at $\mathbf{X}_k$

$$\boldsymbol{\eta}_k = (\mathbf{I}_N - \mathbf{X}_k \mathbf{X}_k^\mathsf{T}) \nabla f_{\mathbf{X}_k}.$$

Now we verify that $\boldsymbol{\eta}_k$ is the gradient of f on $\mathrm{Gr}_{N,s}$ as defined in Definition 2. For any $\boldsymbol{\xi} \in \mathcal{T}_{\mathbf{X}_k} \mathrm{Gr}_{N,s}$,

$$\begin{aligned} \langle \boldsymbol{\eta}_k, \boldsymbol{\xi} \rangle_{\mathbf{X}_k} &= \mathrm{trace}(\nabla f_{\mathbf{X}_k}^\mathsf{T} (\mathbf{I}_N - \mathbf{X}_k \mathbf{X}_k^\mathsf{T}) \boldsymbol{\xi}) \\ &= \mathrm{trace}(\nabla f_{\mathbf{X}_k}^\mathsf{T} \boldsymbol{\xi}) \\ &= \nabla f_{\mathbf{X}_k}[\boldsymbol{\xi}], \end{aligned}$$

in which the second equation comes from that $\mathbf{X}_k^\mathsf{T} \boldsymbol{\xi} = 0$ for all $\boldsymbol{\xi} \in \mathcal{T}_{\mathbf{X}_k} \mathrm{Gr}_{N,s}$.

In the third step of Table I, the retraction of $-t_k \boldsymbol{\eta}_k$ at $\mathbf{X}_k$ is the Q factor of the QR decomposition of $\mathbf{X}_k - t_k \boldsymbol{\eta}_k$. The Armijo step size is $t_k = \alpha \beta^m$ with m being the smallest nonnegative integer such that

$$f(\mathbf{X}_k) - f(R_{\mathbf{X}_k}(-\beta^m \alpha \boldsymbol{\eta}_k)) \geq \sigma \langle \nabla f_{\mathbf{X}_k}, \beta^m \alpha \boldsymbol{\eta}_k \rangle_{\mathbf{X}_k}.$$

Generally, gradient descent line-search algorithm on a compact manifold with Armijo step-size is only guaranteed that the gradient converges. For the subspace projection matrix sensing model (7) and the completion model (12), we can have the following stronger result on the convergence behavior.

Theorem 3. *Suppose that* $\mathbf{y} = \mathcal{A}(\mathbf{P}_{\mathbf{X}_0})$*. For* $\mathcal{A}$ *defined either as* (7) *with the restricted isometry property in Theorem 1, or as* (12)*, every limit point of the sequence* $\{\mathbf{X}_k\}_{k=1}^{\infty}$ *is a critical point of the cost function.*

Proof. The proof is postponed to Appendix C. □

Generally, for a non-convex optimization algorithm relied on the gradient, Theorem 3 can hardly be improved, in that if the iteration starts at a critical point, then the gradient is constantly zero, and the iterates will not move. Therefore, critical point is the best guarantee we can have. Whether or not the iterates converge to a global minimum or a local minimum should be determined by the distribution and property of the critical points of the cost function, as well as the initial point.

TABLE I

GRASSMANN MANIFOLD GRADIENT DESCENT LINE SEARCH ALGORITHM (GGDLS)

Require: continuously differentiable cost function f on $\mathrm{Gr}_{N,s}$;

retraction R from $\mathcal{T}\mathrm{Gr}_{N,s}$ to $\mathrm{Gr}_{N,s}$;

scalars for the Armijo step size $\alpha > 0, \beta, \sigma \in (0,1)$;

Input: Initial iterate $\mathbf{X}_0 \in \mathrm{Gr}_{N,s}$;

Output: Sequence of iterates $\{\mathbf{X}_k\}$.

For $k = 0, 1, 2, \cdots$ **do:**

1.Compute the Euclidean gradient $\nabla f_{\mathbf{X}_k}$ at $\mathbf{X}_k$;

2.Project $\nabla f_{\mathbf{X}_k}$ onto the tangent space $T_{\mathbf{X}_k}\mathrm{Gr}_{N,s}$

to obtain $_k$;

3.$\mathbf{X}_{k+1} = R_{\mathbf{X}_k}(-t_k \;\; _k)$, where t_k is the Armijo step size;

Until: Stopping criterion satisfied;

For subspace projection matrix sensing problem, we can further improve the convergence guarantee of GGDLS in the noiseless scenario, due to the fact that if the RIP condition holds, the only critical point of the cost function in (4) is $\mathbf{X}_0$. Such result is stated as follows.

Proposition 2. *Suppose that* $\mathbf{y} = \mathcal{A}(\mathbf{P}_{\mathbf{X}_0})$*, and* $\mathcal{A}$ *is defined as* (7) *with the restricted isometry property in Theorem 1. If* $\mathbf{X}_k$ *is the result from the* k*-th iteration of the algorithm, then with probability 1, the only limit point of the iterate sequence* $\{\mathbf{X}_k\}_{k=1}^{\infty}$ *is* $\mathbf{X}_0$.

Proof. The proof is postponed to Appendix D. □

VI. NUMERICAL EXPERIMENT

In this section, the proposed algorithm GGDLS will be implemented to retrieve subspace projection matrices from both the two kinds of incomplete information, i.e., the random compressed measurements and the random down-samplings, in both the noiseless and the noise scenarios.

Note that the probability of successful recovery in the following is obtained by counting the times of successful recovery from independent random trails, and a successful recovery means that $\|\mathbf{P}_{\hat{\mathbf{X}}} -$

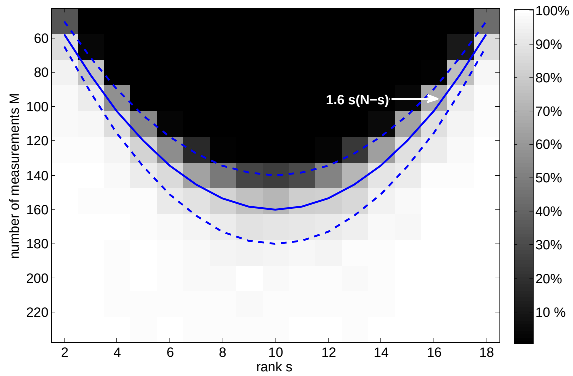

Fig. 1. Successful recovery probability for subspace projection matrix sensing. $N = 20, \alpha = 300, \beta = 0.8, \sigma = 0.01$, random initial value, and 200 trials per point. The dashed lines denote $1.4s(N - s)$ and $1.8s(N - s)$, respectively.

$\mathbf{P}_{\mathbf{X}_0}\|_F/\sqrt{s} \leq 10^{-2}$.

A. Subspace projection matrix sensing

In this part, the performance of GGDLS in the subspace projection matrix sensing problem is demonstrated. The mapping $\mathcal{A}$ is defined as (7), and the corresponding $\mathbf{y}$ in (3) is equivalent to

$$\mathbf{y} = \mathbf{A} \cdot \mathrm{vec}(\mathbf{P}_{\mathbf{X}_0}) + \mathbf{w},$$

in which $\mathbf{A} \in \mathbb{R}^{N^2 \times M}$ is a random Gaussian sensing matrix. The parameters in the Armijo method in GGDLS are $\alpha = 300$, $\beta = 0.8$, and $\sigma = 0.01$. The number of iterations for GGDLS is 200. In each trial, the entries of $\mathbf{A}$ are independently drawn from Gaussian distribution $\mathcal{N}(0, 1/M)$.

In the first experiment, the relation between the required number of measurements M and the rank of the matrix s is shown in the noiseless scenario. We choose $N = 20$ and random initial values. The phase transition map is in Fig. 1, in which the value of each pixel is the probability of successful recovery obtained from 200 trials. The result in Fig. 1 agrees with the theoretical conclusion that $M \gtrsim O(s(N-s))$ is required to guarantee successful recovery with high probability, and here the bound is about $1.6s(N-s)$. It also shows that the algorithm GGDLS is efficient enough to approach the level of the degrees of freedom of subspace projection matrices.

In the second experiment, we test the proposed GGDLS for matrix sensing in the noise scenario. $N = 50$, $s = 5$, and the measurement SNR (MSNR) is varied from 20dB to 60dB with step-size 10dB. The algorithm is initialized by $\mathrm{qr}(\mathrm{reshape}(\mathbf{A}^{\dagger}\mathbf{y}))$. The result is in Fig. 2, in which every point is the

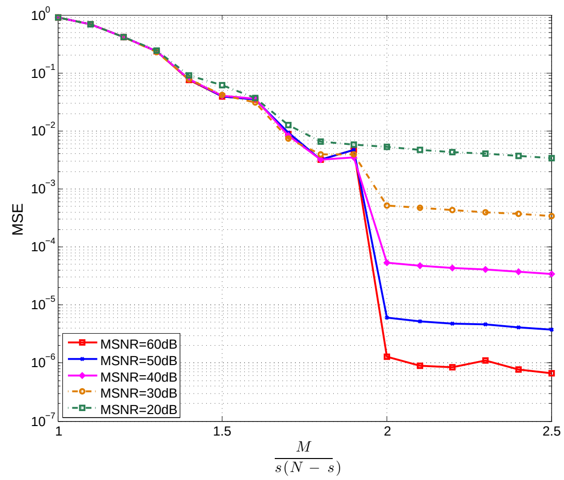

Fig. 2. Recovery performance of subspace projection matrix sensing in noise scenario. $N = 50, s = 5, \alpha = 300, \beta = 0.8, \sigma = 0.01$ and 200 trials per point.

mean square error (MSE) from 200 trials. From Fig. 2, we can see that when $M/(s(N-s))$ is relatively small (less than 2), the MSE is almost the same for various MSNR. However, when $M/(s(N-s))$ is large enough, the logarithm of MSE becomes proportional to the MSNR. Such phenomenon is consistent with the theoretical result that when M is large enough, the RIP holds, so the solution to problem (4) is robust against noise.

B. Subspace projection matrix completion

In this part, we recover the subspace projection matrix from partial entries obtained according to the Bernoulli model (12). The parameters for the Armijo method in GGDLS are $\alpha = 20, \beta = 0.8, \sigma = 0.01$, and the initial values are obtained from the QR decomposition of the incomplete matrix. We will not only demonstrate the performance of the proposed GGDLS, but also compare it with some other efficient low rank matrix completion algorithms.

To begin with, we verify the relation between ρ needed for successful completion with high probability and the rank s in the noiseless case. The size N is set to be 20, and the number of iterations for GGDLS is 1000. The phase transition map is in Fig. 3, in which each point is obtained from 100 random trials. The simulation result in Fig. 3 shows that for successful completion with high probability, ρN^2 should be about $1.8s(N-s)$, which is of the magnitude of the degrees of freedom of subspace projection

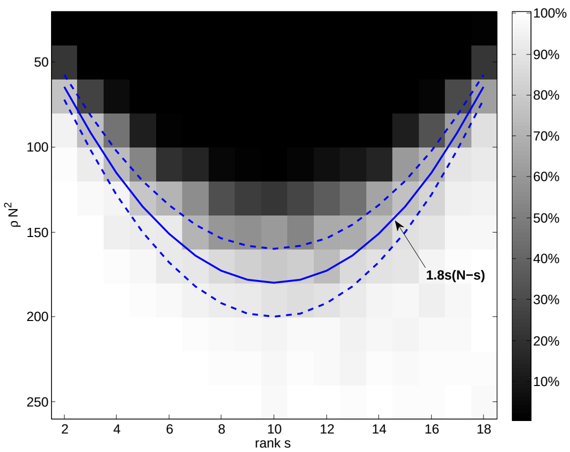

Fig. 3. Successful recovery probability for subspace projection matrix completion. $N = 20, \alpha = 20, \beta = 0.8, \sigma = 0.01$, and 100 trials per point. The dashed lines denote $1.6s(N - s)$ and $2.0s(N - s)$, respectively.

matrices. Note that in the theorem, a regularity condition is needed, while in the experiment there is no such condition, and all of the target matrices are randomly generated.

Next, we compare the proposed GGDLS with some other low rank matrix completion algorithms, which are IALM [41], LMaFit [42], FPCA [43], SRF [44], and OptSpace [45]. These algorithms are designed for general low rank matrices, so for a better comparison in this subspace projection matrix problem, we symmetrize the observed entries because of the precondition that the target matrix is symmetric.

In the fourth experiment, we focus on the sampling density ρ and the probability of successful recovery in the noiseless circumstance. $N = 100, s = 10$, and the number of iterations in GGDLS is set to be 2000. The probability of successful recovery is counted from 100 trails, and the result is in Fig. 4. From the comparison, we can see that the sampling density ρ needed by GGDLS for successful completion with probability 1 is among the lowest, so it is among the most efficient algorithms in terms of sampling density.

In the last experiment, we aim at testing the robustness of these algorithms. $N = 100, s = 10, \rho = 3.5s(N - s)/N^2$, and the number of iterations in GGDLS is set to be 1000. The averaged reconstruction SNR from 100 trials is in Fig. 5. The simulation result shows that GGDLS is the most robust one among all the tested algorithms. FPCA and OptSpace have almost the same behavior, and GGDLS is about 3dB ahead of them.

It should be mentioned that the proposed GGDLS outperforms the reference algorithms due to that it

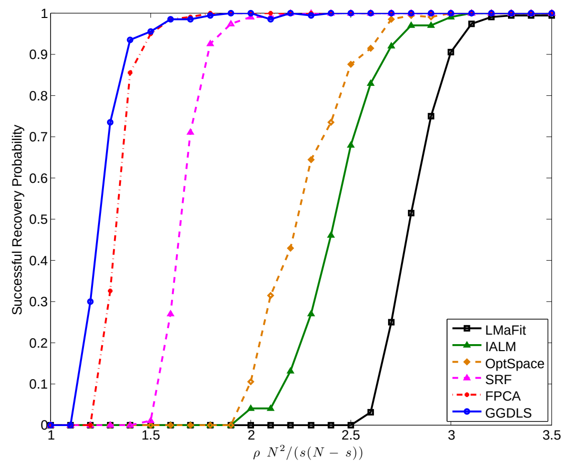

Fig. 4. Successful recovery probability for subspace projection matrix completion by various algorithms. $N = 100, s = 10, \alpha = 20, \beta = 0.8, \sigma = 0.01$.

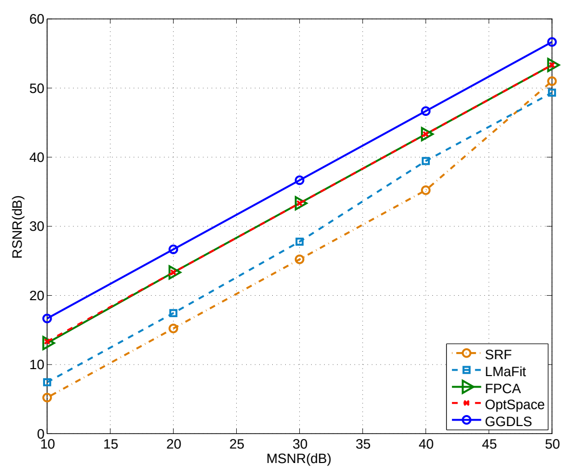

Fig. 5. Recovery performance of subspace projection matrix completion by various algorithms in the noise scenario. $N = 100, s = 10, \rho = 3.5s(N-s)/N^2, \alpha = 20, \beta = 0.8, \sigma = 0.01$.

takes advantage of the specific structure of subspace projection matrices.

VII. Conclusion

In this work, we study the recovery of subspace projection matrix from highly incomplete information in both noiseless and noise cases. The incomplete information that we consider is either from random compressed measurements, i.e., matrix sensing, or from random down-samplings, i.e., matrix completion. An optimization problem on the Grassmann manifold is proposed in the formulation.

For the subspace projection matrix sensing problem, we derive a bound on the number of Gaussian measurements $O(s(N-s))$, so that a restricted isometry property can hold with high probability. The bound is tight in the sense that the dimension of freedom of subspace projection matrix is $s(N-s)$. Such RIP condition can guarantee the unique recovery in the noiseless scenario, and the robust recovery in the noise case.

As for the subspace projection matrix completion problem, we obtain a bound on the sampling density of the Bernoulli model $O\left(s^{3/2}(N-s)(\log N)^3/N^2\right)$ to guarantee an RIP-like inequality for subspace projection matrices with $(\sqrt{2s}\log N)$-regular difference.

A gradient descent algorithm on the Grassmann manifold is proposed to solve the mentioned optimization problem. The convergence behavior of such non-convex optimization algorithm is theoretical analyzed in the noiseless scenario. For the matrix sensing problem, given that the RIP condition holds, the limit point of the iterate sequence will be $\mathbf{X}_0$ with probability 1, and for the matrix completion problem, any limit point will be a critical point of the cost function.

In the numerical experiments, the proposed algorithm is implemented for both the subspace projection matrix sensing problem and the completion problem. The simulation results do not only verify the theoretical results, but also indicate that for the subspace projection matrix completion problem without noise, the bound on the Bernoulli sampling density can reach $O\left(s(N-s)/N^2\right)$. Furthermore, the proposed algorithm is compared with some other low rank matrix completion algorithms both in the noise and the noiseless cases.

This paper is not exhausted, and more works could be done in related topics. For example, the convergence behavior in the noisy circumstances could be theoretically analyzed for both the subspace projection matrix sensing problem and the completion problem. Furthermore, one may try to recover the column space of other kinds of structural matrices with ideas in this paper.

Appendix A
Proof of Theorem 1

Proof. The set of chords of a manifold $\mathcal{M}$ is defined as

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

This proof is mainly established upon the ϵ-net of $\mathcal{C}(\mathcal{P}_{N,s})$ and the standard generic chaining method.

A. Covering net and covering number

Define $N_j(\eta)$ to be the $4^{-j}\eta$-net of $\mathcal{P}_{N,s}$, and $N_j'(\eta,\eta')$ to be the union of the $2^{-j}\eta'$-nets of the ℓ_2 balls in $B_j(\eta)$, in which $B_j(\eta)$, defined as

$$B_j(\eta) := \bigcup_{p\in N_j(\eta)} \{p + B_p\}$$

$$B_p := \{\boldsymbol{\xi} \in \mathcal{T}_p\mathcal{P}_{N,s}, \|\boldsymbol{\xi}\|_2 = 1\},$$

is the union of the unit balls along the tangent space of every point $p \in N_j(\eta)$.

Assume that η and η' are determined by δ, and define $\mathcal{N}_j(\delta) := \mathcal{C}(N_j(\eta)) \bigcup N_j'(\eta,\eta')$, then

$$|\mathcal{N}_j(\delta)| \le \frac{1}{2}|N_j(\eta)|^2 + |N_j(\eta)|\left(\frac{3\cdot 2^j}{\eta'}\right)^{s(N-s)}. \tag{15}$$

According to Lemma 15 in [22], if $\eta = 0.16\tau\delta^2$, $\eta' = 0.4(1.7-\sqrt{2})\delta$, and $\delta < 1/2$, then $\mathcal{N}_j(\delta)$ is a $2^{-j}\delta$-net for $\mathcal{C}(\mathcal{P}_{N,s})$. So we have found a covering net.

As for the covering number $|\mathcal{N}_j(\delta)|$, remind that the projection distance on $\mathrm{Gr}_{N,s}$ is

$$d_p^2(\mathbf{X},\mathbf{Y}) := \|\mathbf{X}\mathbf{X}^\mathsf{T} - \mathbf{Y}\mathbf{Y}^\mathsf{T}\|_F^2/2,$$

so the $\sqrt{2}\delta$ covering number of $\mathcal{P}_{N,s}$ is the δ covering number of $\mathrm{Gr}_{N,s}$ with respect to the projection distance, which is bounded above by $(C_0/\delta)^{s(N-s)}$ according to [46].

Therefore, by plugging the above in (15), we have

$$\begin{aligned}|\mathcal{N}_j(\delta)| \le \frac{1}{2}&\left(\frac{C_0 4^j}{0.16\tau\delta^2}\right)^{2s(N-s)} + \\ &\left(\frac{3\cdot 2^j}{0.4(1.7-\sqrt{2})\delta}\right)^{s(N-s)}\left(\frac{C_0 4^j}{0.16\tau\delta^2}\right)^{s(N-s)}.\end{aligned}$$

From Lemma 1, we know that $\tau = 1/\sqrt{2}$ for $\mathcal{P}_{N,s}$. Denote $\bar{C}_0 = \max\left\{\frac{\sqrt{2}C_0}{0.16}, \frac{3}{0.4(1.7-\sqrt{2})}\right\} > 1$, then

$$\begin{aligned}|\mathcal{N}_j(\delta)| &\le \frac{1}{2}\left(\frac{\bar{C}_0 4^j}{\delta^2}\right)^{2s(N-s)} + \\ &\quad\left(\frac{\bar{C}_0\cdot 2^j}{\delta}\right)^{s(N-s)}\left(\frac{\bar{C}_0 4^j}{\delta^2}\right)^{s(N-s)} \\ &\le \left(\frac{\bar{C}_0 4^j}{\delta^2}\right)^{2s(N-s)}.\end{aligned}$$

B. Generic chaining

Notice that $\forall \mathbf{Z} \in \mathcal{C}(\mathcal{P}_{N,s})$ we can decompose it as

$$\mathbf{Z} = \pi_0 + \sum_{j=0}^{\infty} (\pi_{j+1}(\mathbf{Z}) - \pi_j(\mathbf{Z})), \tag{16}$$

in which $\pi_j(\mathbf{Z})$ is the nearest point to $\mathbf{Z}$ in $\mathcal{N}_j(\delta)$, and

$$\|\pi_{j+1}(\mathbf{Z}) - \pi_j(\mathbf{Z})\|_2 \leq 2^{-j+1}\delta.$$

According to the triangle inequality, we know that for fixed $0 < \delta < \epsilon_1 < \epsilon \leq \frac{1}{3}$,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\mathbf{Z})\|_2 - 1| > \epsilon\right\} \\
&\leq 2\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\pi_0(\mathbf{Z}))\|_2 - 1| > \epsilon_1\right\} \\
&+ 2\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\Sigma(\mathbf{Z}))\|_2 - 1| > \epsilon - \epsilon_1\right\},
\end{aligned}$$

in which $\Sigma(\mathbf{Z})$ is the sum in (16).

From Lemma 16 and Lemma 17 in [22], we know that for C_5 satisfying $\frac{\epsilon_1}{\delta} \geq \frac{1+C_5}{1-C_5}$ and C_6 satisfying $\frac{\epsilon-\epsilon_1}{\delta} \geq C_6$, if $M \geq 14\epsilon_1^{-2}\log(|\mathcal{N}_0(\delta)|)$, then

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\pi_0(\mathbf{Z}))\|_2 - 1| > \epsilon_1\right\} \\
&\leq |\mathcal{N}_0(\delta)| \cdot \max_{t_0\in\mathcal{N}_0(\delta)} \mathbb{P}\left\{\frac{|\|\mathcal{A}(t_0)\|_2 - \|t_0\|_2|}{\|t_0\|_2} > C_5\epsilon_1\right\} \\
&\leq 2\exp(-M\epsilon_1^2/14),
\end{aligned}$$

and if $M \geq 14\epsilon_1^{-2}\log(4|\mathcal{N}_0(\delta)|)$, then

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\Sigma(\mathbf{Z}))\|_2 - 1| > \epsilon - \epsilon_1\right\} \\
&\leq \sum_j |\mathcal{N}_{j+1}(\delta)|^2 \cdot \\
&\quad \max_{\substack{(t_j,s_j)\in\mathcal{N}_j(\delta)\\ \times\mathcal{N}_{j+1}(\delta)}} \mathbb{P}\left\{\frac{\|\mathcal{A}(t_j - s_j)\|_2}{\|t_j - s_j\|_2} > \frac{C_6(j+1)}{8}\right\} \\
&\leq \sum_j |\mathcal{N}_{j+1}(\delta)|^2 \mathrm{e}^{-(2j+1)M/7} \\
&\leq 2\exp(-M(1-\epsilon_1^2)/7) \\
&\leq 2\exp(-M\epsilon_1^2/14).
\end{aligned}$$

Therefore, we have the conclusion that

$$\mathbb{P}\left\{\sup_{\mathbf{Z}\in\mathcal{C}(\mathcal{M})} |\|\mathcal{A}(\mathbf{Z})\|_2 - 1| > \epsilon\right\} \leq \rho$$

if $M \geq 14\epsilon_1^{-2} \cdot \max\{2s(N-s)\log(4\bar{C}_0/\delta^2), \log(8/\rho)\}$. By taking $\epsilon_1 = \frac{9}{10}\epsilon$ and $\delta = \frac{\epsilon}{160}$, denote $c_0 = 4 \times 160^2 \bar{C}_0$, we have arrived at (8). □

APPENDIX B
PROOF OF THEOREM 2

Proof. For any $\mathbf{P}_\mathbf{X}$ and $\mathbf{P}_\mathbf{Y}$ satisfying the condition in Theorem 2, denote $\mathbf{Z} := \frac{\mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{Y}}{\|\mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{Y}\|_F}$, then $\mathbf{Z} \in \mathcal{C}(\mathcal{P}_{N,s})$ and $\|\mathbf{Z}\|_\infty \leq \frac{\alpha}{N}$. Define a set $\mathcal{C}_\alpha := \{\mathbf{Z} \in \mathcal{C}(\mathcal{P}_{N,s}) : \|\mathbf{Z}\|_\infty \leq \frac{\alpha}{N}\}$. The proof is established on the covering number of $\mathcal{C}_\alpha$, the Bernstein's inequality, and the union bound.

First, denote $\mathcal{N}(\mathcal{P}_{N,s}, T)$ as the T-covering of $\mathcal{P}_{N,s}$, and define B_T as

$$B_T := \bigcup_{p\in\mathcal{N}(\mathcal{P}_{N,s},T)} \{p + B_{p,T}\}$$

$$B_{p,T} := \{\boldsymbol{\xi} \in \mathcal{T}_p\mathcal{P}_{N,s}, \|\boldsymbol{\xi}\|_F \leq T\}.$$

According to [21], $\mathcal{C}(B_T)$ is a $4\sqrt{T/\tau}$ covering of $\mathcal{C}(\mathcal{P}_{N,s})$. (The notation $\mathcal{C}(\cdot)$ is defined in Appendix A.) Thus, the ϵ covering of $\mathcal{C}(B_T)$, denoted as $\mathcal{N}(\mathcal{C}(B_T), \epsilon)$, is an $(\epsilon + 4\sqrt{T/\tau})$-covering of $\mathcal{C}(\mathcal{P}_{N,s})$. Therefore, we can construct a $2(\epsilon + 4\sqrt{T/\tau})$-covering set $\mathcal{N}(\mathcal{C}_\alpha, \epsilon, T)$ for $\mathcal{C}_\alpha$, such that $|\mathcal{N}(\mathcal{C}_\alpha, \epsilon, T)| < |\mathcal{N}(\mathcal{C}(B_T), \epsilon)|$, and $\mathcal{N}(\mathcal{C}_\alpha, \epsilon, T) \subset \mathcal{C}_\alpha$. From [21] we know that $|\mathcal{N}(\mathcal{C}(B_T), \epsilon)| \leq (\sqrt{2}C_0/T)^{s(N-s)}(1 +$

$2/\epsilon)^s + (\sqrt{2}C_0/T)^{2s(N-s)}(1+2/\epsilon)^{2s+1}$. So we have obtained a bound on the covering number. Next we need to bound the probability.

According to Lemma 2.5 in [29], which is based on the Bernstein's inequality, we know that if $\hat{\mathbf{Z}}$ is α-regular, then

$$\mathbb{P}\left\{|\|\mathcal{A}(\hat{\mathbf{Z}})\|_F^2 - \rho\|\hat{\mathbf{Z}}\|_F^2| > \delta\rho\|\hat{\mathbf{Z}}\|_F^2\right\} \le 2\exp\left(-\frac{\delta^2\rho N^2}{3\alpha}\right).$$

Based on the union bound, we have that

$$\begin{aligned}&\mathbb{P}\left\{\forall\hat{\mathbf{Z}} \in \mathcal{N}(\mathcal{C}_\alpha, \epsilon, T), |\|\mathcal{A}(\hat{\mathbf{Z}})\|_F^2 - \rho\|\hat{\mathbf{Z}}\|_F^2| > \delta\rho\|\hat{\mathbf{Z}}\|_F^2\right\}\\ &\le 2|\mathcal{N}(\mathcal{C}_\alpha, \epsilon, T)|\exp\left(-\frac{\delta^2\rho N^2}{3\alpha}\right).\end{aligned}$$

Since $\|\hat{\mathbf{Z}}\|_F = 1$ and $\|\mathbf{Z}\|_F = 1$, we are able to control the bound by two parts as the following

$$\begin{aligned}\sup_{\mathbf{Z}\in\mathcal{C}_\alpha}\|\mathcal{A}(\mathbf{Z})\|_2 &\le \sup_{\substack{\mathbf{Z}\in\mathcal{C}_\alpha,\\ \hat{\mathbf{Z}}\in\mathcal{N}(\mathcal{C}_\alpha,\epsilon,T)}} \|\mathcal{A}(\mathbf{Z}-\hat{\mathbf{Z}})\|_2 + \|\mathcal{A}(\hat{\mathbf{Z}})\|_F\\ &\le 2(\epsilon + 4\sqrt{T/\tau}) + \sqrt{(1+\delta)\rho} \qquad (17)\end{aligned}$$

$$\begin{aligned}\inf_{\mathbf{Z}\in\mathcal{C}_\alpha}\|\mathcal{A}(\mathbf{Z})\|_2 &\ge \inf_{\substack{\mathbf{Z}\in\mathcal{C}_\alpha,\\ \hat{\mathbf{Z}}\in\mathcal{N}(\mathcal{C}_\alpha,\epsilon,T)}} \|\mathcal{A}(\hat{\mathbf{Z}})\|_2 - \|\mathcal{A}(\mathbf{Z}-\hat{\mathbf{Z}})\|_F\\ &\ge \sqrt{(1-\delta)\rho} - 2(\epsilon + 4\sqrt{T/\tau}). \qquad (18)\end{aligned}$$

The last two inequalities hold with probability no less than $1 - 2|\mathcal{N}(\mathcal{C}_\alpha, \epsilon, T)|\exp\left(-\frac{\delta^2\rho N^2}{3\alpha}\right)$.

Set $\delta = \varepsilon/2$, $\epsilon = \sqrt{\rho\varepsilon}/(4\sqrt{2})$, and $T = \tau\rho\varepsilon/512$. If ρ satisfies

$$\begin{aligned}\rho \ge &\frac{12\alpha\beta}{N^2\varepsilon^2}\Bigg\{\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right)\\ &+ 2s(N-s)\cdot\max\left[0, \log\left(\frac{1024C_0}{\varepsilon\rho}\right)\right]\Bigg\}, \qquad (19)\end{aligned}$$

then (14) can be achieved with probability no less than $1-2\mathrm{e}^{-\beta}$. Denote a constant $C_1 := \max[0, \log(1024C_0)]$, then (19) is equivalent to

$$\begin{aligned}\rho \ge &\frac{12\alpha\beta}{N^2\varepsilon^2}\Bigg[\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right)\\ &+ 2C_1 s(N-s)\log\left(\frac{1}{\varepsilon\rho}\right)\Bigg]. \qquad (20)\end{aligned}$$

A necessary condition for (20) to hold is that

$$\rho \geq \frac{12\alpha\beta}{N^2\varepsilon^2}\left[\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right)\right.$$
$$\left. + 2C_1 s(N-s)\log\left(\frac{1}{\varepsilon}\right)\right],$$

which gives an upper bound on $\log(1/\rho)$

$$\log\frac{1}{\rho} \leq -\log\left\{\frac{12\alpha\beta}{N^2\varepsilon^2}\left[\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right)\right.\right.$$
$$\left.\left. + 2C_1 s(N-s)\log\left(\frac{1}{\varepsilon}\right)\right]\right\}$$
$$\leq -\log\left(\frac{12\alpha\beta}{N^2\varepsilon^2}\right).$$

Therefore, we know that a sufficient condition for (20) to be satisfied is that

$$\rho \geq \frac{12\alpha\beta}{N^2\varepsilon^2}\left\{\log 2 + (2s+1)\log\left(\frac{4}{\varepsilon}\right)\right.$$
$$\left. + 2C_1 s(N-s)\left[\log\left(\frac{1}{\varepsilon}\right) + \log\left(\frac{N^2\varepsilon^2}{12\alpha\beta}\right)\right]\right\},$$

which is exactly (13). □

APPENDIX C
PROOF OF THEOREM 3

Proof. Denote $\mathbf{g}_k := \mathrm{grad} f(\mathbf{X}_k)$. Since $\mathrm{Gr}_{N,s}$ is a compact manifold, it is known that [37]

$$\|\mathbf{g}_k\|_2 \to 0. \tag{21}$$

By definition $\langle \mathbf{g}_k, \boldsymbol{\xi}\rangle = \nabla f(\mathbf{X}_k)[\boldsymbol{\xi}], \forall \boldsymbol{\xi} \in \mathcal{T}_{\mathbf{X}_k}\mathcal{M}$, we have

$$\forall \boldsymbol{\xi}_k \in \mathcal{T}_{\mathbf{X}_k}\mathcal{M}, \lim_{k\to\infty}\langle \mathbf{g}_k, \boldsymbol{\xi}_k\rangle = \lim_{k\to\infty}\nabla f(\mathbf{X}_k)[\boldsymbol{\xi}_k] = 0.$$

Therefore, $\nabla f(\mathbf{X}_k)$ tends to the zero matrix as k tends to ∞.

Suppose that $\mathbf{X}^0$ is a limit point of the sequence $\{\mathbf{X}_k\}$, i.e., $\forall \delta > 0$ there exists an infinite subsequence of $\{\mathbf{X}_k\}$ in $B(\mathbf{X}^0, \delta)$ which converges to $\mathbf{X}^0$. If $\mathbf{X}^0$ is not a critical point, then $\exists a, b$ such that $[\nabla f(\mathbf{X}^0)]_{a,b} \neq 0$. Without loss of generality, we can assume that $[\nabla f(\mathbf{X}^0)]_{a,b} > 0$, then there exists $\varepsilon > 0$ such that $[\nabla f(\mathbf{X}^0)]_{a,b} > \varepsilon$.

For a fixed

$$\delta = \frac{\|\widetilde{\mathbf{A}_a}\|}{2(\|\mathcal{A}(\mathbf{P}_{\mathbf{X}^0} - \mathbf{P}_{\mathbf{X}_0})\|_2 + 1 + \epsilon)}, \tag{22}$$

for all $\mathbf{X} \in \mathrm{Gr}_{N,s}$ such that $d_p(\mathbf{X}, \mathbf{X}^0) = \|\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}^0}\|_F \leq \delta$,

$$\begin{aligned}
[\nabla f(\mathbf{X})]_{a,b} &= \mathcal{A}(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_0})^{\mathsf{T}} \cdot \widetilde{\mathbf{A}_a} \cdot \mathbf{x}_b \\
&= \mathcal{A}(\mathbf{P}_{\mathbf{X}^0} - \mathbf{P}_{\mathbf{X}_0})^{\mathsf{T}} \cdot \widetilde{\mathbf{A}_a} \cdot \mathbf{x}_b^0 \\
&\quad + \mathcal{A}(\mathbf{P}_{\mathbf{X}^0} - \mathbf{P}_{\mathbf{X}_0})^{\mathsf{T}} \cdot \widetilde{\mathbf{A}_a} \cdot (\mathbf{x}_b - \mathbf{x}_b^0) \\
&\quad + \mathcal{A}(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}^0})^{\mathsf{T}} \cdot \widetilde{\mathbf{A}_a} \cdot \mathbf{x}_b \\
&> \varepsilon - \|\mathcal{A}(\mathbf{P}_{\mathbf{X}^0} - \mathbf{P}_{\mathbf{X}_0})\|_2 \|\widetilde{\mathbf{A}_a}\| \|\mathbf{x}_b - \mathbf{x}_b^0\|_2 \\
&\quad - \|\mathcal{A}(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}^0})\|_2 \|\widetilde{\mathbf{A}_a}\| \| \|\mathbf{x}_b\|_2 \\
&\geq \varepsilon - \delta \|\widetilde{\mathbf{A}_a}\| (\|\mathcal{A}(\mathbf{P}_{\mathbf{X}^0} - \mathbf{P}_{\mathbf{X}_0})\|_2 + 1 + \epsilon),
\end{aligned}$$

in which the last inequality holds due to $\|\mathbf{x}_b\|_2 = 1$ and $\|\mathbf{x}_b - \mathbf{x}_b^0\|_2 \leq \|\mathbf{X} - \mathbf{X}^0\|_F \leq \|\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}^0}\|_F$. For matrix sensing ϵ is in (9), and for matrix completion $\epsilon = 0$. By choose δ as (22), we have $[\nabla f(\mathbf{X})]_{a,b} > \varepsilon/2$.

Therefore, $\exists \varepsilon > 0$ and $\delta > 0$, such that the infinite sub-sequence of $\{\mathbf{X}_k\}$ in $B(\mathbf{X}^0, \delta)$ converging to $\mathbf{X}^0$ satisfies $\|\nabla f(\mathbf{X}_k)\|_F > \varepsilon/2$, which contradicts to $\nabla f(\mathbf{X}_k) \to \mathbf{0}$. So we have that if $\mathbf{X}^0$ is a limit point of $\{\mathbf{X}_k\}$, then $\nabla f(\mathbf{X}^0)$ equals to the zeros matrix. □

APPENDIX D
PROOF OF PROPOSITION 2

Proof. From the fact that

$$\begin{aligned}
[\nabla f(\mathbf{X})]_{a,b} &= \sum_{j=1}^{M} \langle \mathbf{A}_j, \mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_0} \rangle [\widetilde{\mathbf{A}_a} \mathbf{x}_b]_j \\
&= \sum_{j=1}^{M} \sum_{i=1}^{N} \langle \mathbf{A}_j, \mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_0} \rangle ([\mathbf{A}_j]_{i,a} + [\mathbf{A}_j]_{a,i}) [\mathbf{X}]_{i,b},
\end{aligned}$$

we obtain that

$$\nabla f(\mathbf{X}) = \sum_{j=1}^{M} \langle \mathbf{A}_j, \mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{X}_0} \rangle (\mathbf{A}_j + \mathbf{A}_j^{\mathsf{T}}) \cdot \mathbf{X}.$$

Suppose that $\mathbf{X}^*$ is a critical point, i.e., $\nabla f(\mathbf{X}^*)$ equals to the zero matrix. A necessary condition is that the rank of the $N \times N$ matrix $\sum_{j=1}^{M}\langle \mathbf{A}_j, \mathbf{P}_{\mathbf{X}^*} - \mathbf{P}_{\mathbf{X}_0}\rangle(\mathbf{A}_j + \mathbf{A}_j^\mathsf{T})$ has to be no larger than $N - s$, so

$$\det\left(\sum_{j=1}^{M}\langle \mathbf{A}_j, \mathbf{P}_{\mathbf{X}^*} - \mathbf{P}_{\mathbf{X}_0}\rangle(\mathbf{A}_j + \mathbf{A}_j^\mathsf{T})\right) = 0.$$

For any fixed $\mathbf{P}_{\mathbf{X}^*} \neq \mathbf{P}_{\mathbf{X}_0}$, we claim that such condition holds with probability 0. In order to prove this claim, we only need to see that the zero set of the above determinant function in each entry of every $\mathbf{A}_j$ is of measure zero, in that $\{\mathbf{A}_j\}_{j=1}^{M}$ are independent Gaussian matrices with independent entries. In fact, if a real analytic function is zero on a set of positive measure, then it is identically zero. Since the determinant function is analytic in each $[\mathbf{A}_j]_{k,l}$, and the function is not identically zero, we have shown that the claim is correct.

According to Theorem 3, all the limit points of the sequence $\{\mathbf{X}_k\}$ are critical points. Since now with probability 1 the only critical point is $\mathbf{X}_0$, we have that with probability 1 the only limit point is $\mathbf{X}_0$. $\square$

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC 61371137). The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

REFERENCES

[1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing mri. volume 25, pages 72–82. IEEE, 2008.

[2] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009.

[3] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19:2861–2873, November 2010.

[4] G. E. Hinton, P. Dayan, and M. Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Netw.*, 8(1):65C74, 1997.

[5] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2004.

[6] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[7] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[8] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[9] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009.

[10] W. Maetzel, J. Romberg, K. G. Sabra, and W. Kuperman. Compressive matched field processing. *The Journal of the Acoustical Society of America*, 127:1908, 2010.

[11] W. Mantzel. *Parametric estimation of randomly compressed functions*. PhD thesis, Georgia Institute of Technology, 2013.

[12] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.

[13] A. M. So and Y. Ye. Theory of semidefinite programming for sensor network localization. In *IN SODA05*, pages 405–414, 2005.

[14] J. Bennett, S. Lanning, and Netflix. The netflix prize. In *In KDD Cup and Workshop in conjunction with KDD*, 2007.

[15] E. J. Candès and B. Recht. Exact low-rank matrix completion via convex optimization. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 806–812, Sept 2008.

[16] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th annual ACM symposium on Theory of computing, 2013*, pages 665–674.

[17] E. J. Candès and Y. Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *Information Theory, IEEE Transactions on*, 57(4):2342–2359, April 2011.

[18] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[19] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[20] X. Shen and Y. Gu. Restricted isometry property of subspace projection matrix under random compression. *Signal Processing Letters, IEEE*, 22(9):1326–1330, Sept 2015.

[21] H. L. Yap, M. B. Wakin, and C. J. Rozell. Stable manifold embeddings with structured random matrices. *Selected Topics in Signal Processing, IEEE Journal of*, 7(4):720–730, Aug 2013.

[22] A. Eftekhari and M. B. Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *CoRR*, abs/1306.4748, 2013.

[23] M. A. Iwen and M. Maggioni. Approximation of points on low-dimensional manifolds via random linear projections. *CoRR*, abs/1204.3337, 2012.

[24] K. L. Clarkson. Tighter bounds for random projections of manifolds. In *Symposium on Computational Geometry*, pages 39–48. ACM, 2008.

[25] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere. preprint,*http://arxiv.org/abs/1504.06785*, 2015.

[26] S. Mittal and P. Meer. Conjugate gradient on grassmann manifolds for robust subspace estimation. *Image and Vision Computing*, 30(6-7):417 – 427, 2012.

[27] L. Balzano, R. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 704–711, Sept 2010.

[28] B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

[29] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *In NIPS 2010*, 2010.

[30] K. C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting, May 2005.

[31] S. Rao, R. Tron, R. Vidal, and Y. Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, Oct 2010.

[32] J. Ho, K. C. Lee, M. H. Yang, and D. Kriegman. Visual tracking using learned linear subspaces. In *Computer Vision and Pattern Recognition, 2004 IEEE Computer Society Conference on*, volume 1, pages I–782–I–789 Vol.1, June 2004.

[33] L. Balzano, R. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 704–711, Sept 2010.

[34] Y. Chi, Y. C. Eldar, and R. Calderbank. Petrels: Subspace estimation and tracking from partial observations. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012.

[35] X. Mao and Y. Gu. Compressed subspace clustering: A case study. In *Global Conference on Signal and Information Processing, 2014 IEEE*, Dec 2014.

[36] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39(1-3):419–441, 2008.

[37] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[38] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589 – 592, 2008.

[39] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. 2008.

[40] X. Shen and Y. Gu. Subspace projection matrix completion on grassmann manifold. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, April 2015.

[41] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2013.

[42] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4(4):333–361, 2012.

[43] S. Ma, D. Goldfarb, and L. Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011.

[44] H. Ghasemi, M. Malek-Mohammadi, M. Babaie-Zade, and C. Jutten. Srf: Matrix completion based on smoothed rank function. In *Proceedings of ICASSP 2011*, pages 3672–3675.

[45] R. H. Keshavan and S. Oh. A gradient descent algorithm on the grassman manifold for matrix completion. Sep 2009.

[46] S. J. Szarek. Metric entropy of homogeneous spaces. In *in Quantum Probability, Banach Center Publ. 43, Polish Acad. Sci*, pages 395–410, 1998.