Low-Rank Matrix Completion with Geometric Performance Guarantees

Wei Dai*, Ely Kerman**, Olgica Milenkovic*
*Department of Electrical and Computer Engineering, **Department of Mathematics
University of Illinois at Urbana-Champaign
Email: {weidai07,ekerman,milenkov}@illinois.edu

Abstract—The low-rank matrix completion problem can be stated as follows: given a subset of the entries of a matrix, find a low-rank matrix consistent with the observations. There exist several low-complexity algorithms for low-rank matrix completion which focus on the minimization of the Frobenius norm of the matrix projection residue. This optimization framework has inherent difficulties: the objective function is not continuous and the solution set is not closed. To address this problem, we propose a geometric objective function to replace the Frobenius norm: the new objective function is continuous everywhere and the solution set is the closure of the solution set of the Frobenius norm. Furthermore, using the geometric objective function and a simple gradient descent procedure, we are able to preclude the existence of local minimizers, and hence establish strong performance guarantees for special completion scenarios, which do not require matrix incoherence or large matrix size.

Index Terms—geometry, low rank, matrix completion.

I. INTRODUCTION

In many practical applications of data acquisition, the signals of interest have a sparse representation in some basis. That is, they can be well approximated using only a few basis elements. This allows for efficient sampling and reconstruction of signals [1], [2], [3], [4]. Two categories of sparse signals are of particular interest: sparse vectors and low-rank matrices. Compressive sensing is the framework of sampling and recovering sparse vectors, while low-rank matrix completion studies when and how one can recover a low-rank matrix based on only a subset of its entries [3], [4].

Scores of methods and algorithms have been proposed for low-rank matrix completion. Many approaches for matrix completion can be viewed as generalizations of their counterparts for compressive sensing reconstruction: the nuclear norm minimization [3], [4] is based on the $\ell_1$-minimization; the ADMiRA algorithm [5] is a counterpart of the SP [6] and CoSaMP [7] algorithms; and the SVP method [8] extends the IHT [9] approach. There are also other approaches that utilize some special structural properties of the low-rank matrices. Examples include the powerfactorization algorithm [10], the OptSpace algorithm [11], and the SET algorithm [12].

Nevertheless, there is a fundamental open problem in low-rank matrix completion: how to perform $\ell_0$-search for low-rank matrices. There is a fundamental difference between compressive sensing and low-rank matrix completion. In compressive sensing, the basis under which the signal is sparse is known a priori. The exhaustive search for a sparse solution is computationally expensive but doable in principle. On the other hand, in low-rank matrix completion, the corresponding “sparse basis” is not known. The set of all possible bases forms a continuous space, in which “exhaustive” search is impossible.

There are two special cases where specialized algorithms can guarantee a low-rank solution consistent with the observations. The first case is when the low-rank matrix is fully sampled. The low-rank solution is simply the observation matrix itself. The corresponding “sparse basis” (singular vectors) can be easily obtained by a singular value decomposition. The other case is when the rank is equal to one. Given an arbitrary sampling pattern, one only has to record the ratios between the revealed entries in the same column and uses these ratios to construct a column vector that represents the column space. Note that neither of these two methods is universal, i.e., neither can be applied for matrices other than the one they are specialized for.

The main contribution of this work is to provide a new direction for low-rank matrix completion, based on geometric performance measures. In particular, we describe a new objective function for which gradient descent algorithm has provable performance guarantees for both rank-one and full-sampling scenarios, independent of the incoherence property or size of the matrices.

The main ideas behind the proposed geometric approach are summarized below:

1) We assume that the goal of the completion algorithm is to find a low-rank matrix that is consistent with the partial observations. Such a matrix may not be unique: if there is a unique low-rank solution, we should be able to find this unique solution; otherwise, it suffices to find at least one solution that agrees with the revealed entries. We assume that the rank of the underlying low-rank matrix is known a priori. Finding a consistent low-rank matrix is equivalent to finding a consistent column/row space. Our consistent completion algorithm differs from the OptSpace algorithms [11], where the search is performed on both column and row spaces simultaneously.

2) We propose a geometric performance metric to measure consistency. In literature, the standard approach for solving the completion problem is to minimize an objective function defined via the Frobenius norm. Unfortunately, as we shall illustrate later, the Frobenius norm leads to a discontinuous objective function and the corresponding solution set is not
closed. To address this issue, we introduce a performance metric based on principle angles which leads to a continuous objective function. The set of the geometrically consistent solutions is the closure of the solution set corresponding to the Frobenius norm. This closure property holds for arbitrary low-rank matrices with arbitrary sampling pattern.

3) We describe a general method for performing “exhaustive search” for consistent low-rank matrices. The core of the method is a randomized gradient descent that uses the geometric objective function. This method is guaranteed to converge to a consistent solution with probability one in two special scenarios: rank-one matrices with arbitrary sampling patterns, and fully sampled matrices of arbitrary rank. More importantly, in these two scenarios, if the observations admit a unique consistent solution, this method finds this unique solution with probability one. The performance guarantees are stronger than those previously established in literature: most known performance guarantees are based on incoherence conditions, they rely on large matrix sizes and only hold with high probability; ours do not require incoherence conditions and hold with probability one, regardless of the matrix size.

II. LOW-RANK MATRIX COMPLETION AND PRELIMINARIES

Let \( X \in \mathbb{R}^{m \times n} \) be an unknown matrix with rank \( r \leq \min(m, n) \), and let \( \Omega \subset [n] \times [n] \) be the set of indices of the observed entries, where \( [k] = \{1, 2, \ldots, k\} \). Define the projection operator \( P_\Omega : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n} \) by

\[
P_\Omega(X) = X_\Omega, \quad \text{where} \quad (X_\Omega)_{i,j} = \begin{cases} X_{i,j} & \text{if } (i,j) \in \Omega \\ 0 & \text{if } (i,j) \notin \Omega. \end{cases}
\]

The consistent matrix completion problem is to find one rank-\( r \) matrix \( X' \) that is consistent with the observations \( X_\Omega \), i.e.,

\[
(P0): \text{find a } X' \text{ such that}\]

\[
\text{rank} (X') = r \quad \text{and} \quad P_\Omega (X') = P_\Omega (X) = X_\Omega. \tag{1}
\]

By definition, this problem is well defined since \( X_\Omega \) is obtained from some rank-\( r \) matrix \( X \) which is therefore a solution. As in other works, \([5],[10],[11]\), we assume that the rank \( r \) is given. In practice, one may try to sequentially guess a rank bound until a satisfactory solution has been found.

We also introduce the (standard) projection operator \( P : \mathbb{R}^m \times \mathbb{R}^{m \times k} \rightarrow \mathbb{R}^m \) given by \( P(x, U) \rightarrow y = UU^Tx \), where \( 1 \leq k \leq m \), and where the superscript \( \dagger \) denotes the pseudoinverse of a matrix. Let \( \text{span}(U) = \{u \in \mathbb{R}^m : v = Uw \text{ for some } w \in \mathbb{R}^m\} \) be the subspace spanned by the columns of the matrix \( U \). One can describe \( P(x, U) \), in geometric terms, as the projection of the vector \( x \) onto \( \text{span}(U) \).

A. Search for a consistent column space

Let \( U_{m,r} \) be the set of \( m \times r \) matrices with \( r \) orthonormal columns, i.e., \( U_{m,r} = \{U \in \mathbb{R}^{m \times r} : U^TU = I_r\} \). Define the function \( f_F : U_{m,r} \rightarrow \mathbb{R} \) by setting

\[
f_F(U) = \min_{W \in \mathbb{R}^{r \times r}} \left\| X_\Omega - P_\Omega(UW^T) \right\|_F^2, \tag{2}
\]

where \( \| \cdot \|_F \) denotes the Frobenius norm. This function measures the consistency between the matrix \( U \) and the observations \( X_\Omega \). In particular, if \( f_F(U) = 0 \), then there exists a matrix \( W \) such that the rank-\( r \) matrix \( UW^T \) satisfies \( P_\Omega(UW^T) = X_\Omega \). Hence, the consistent matrix completion problem is equivalent to

\[
(P1): \text{find } U \in U_{m,r} \text{ such that } f_F(U) = 0. \tag{3}
\]

In fact, \( f_F(U) \) depends only on the column space \( \text{span}(U) \). Hence, to solve the consistent matrix completion problem, it suffices to find a column space that is consistent with the observed entries. Note that the same conclusion holds for the row space as well. For simplicity, we focus our attention to the column space only.

B. Grassmann Manifolds

The Grassmann manifold \( G_{m,r} \) is the set of all \( r \)-dimensional subspaces in the \( m \)-dimensional Euclidean space \( \mathbb{R}^m \). This is a smooth compact manifold of dimension \( r(m - r) \). Every element, say \( \mathcal{W} \in G_{m,r} \), can be presented by a generator matrix \( U \in U_{m,r} \) satisfying \( \text{span}(U) = \mathcal{W} \). Problem (P1) can be viewed as an optimization problem on the compact manifold \( G_{m,r} \).

Principal angles and chordal distance on \( G_{m,r} \). Consider the subspaces \( \text{span}(U) \) and \( \text{span}(V) \) of \( \mathbb{R}^m \) for some \( U, V \in U_{m,r} \). If \( 1 \leq r \leq m \), the principal angles between these two subspaces can be defined as follows. Let \( u_1 \in \text{span}(U) \) and \( v_1 \in \text{span}(V) \) be unit-length vectors such that \( u_1^Tv_1 \) is maximal. Inductively, let \( u_k \in \text{span}(U) \) and \( v_k \in \text{span}(V) \) be unit-length vectors such that \( u_k^Tv_j = 0 \) for \( 1 \leq j < k \) and \( u_k^Tv_k \) is maximal. The principal angles are then defined as \( \alpha_k = \arccos u_k^Tv_k \) for \( k = 1, 2, \ldots, r \). Alternatively, the principal angles can be computed via singular value decomposition. Consider the singular value decomposition \( UV^TVV^T = U\Lambda V^T \), where \( U \in U_{m,r} \) and \( V \in U_{m,r} \) contain the first \( r \) left and right singular vectors, respectively, and \( \Lambda \in \mathbb{R}^{r \times r} \) is a diagonal matrix comprised of singular values \( \lambda_1 \geq \cdots \geq \lambda_r \). Then the \( k \)-th columns of \( U \) and \( V \) correspond to the vectors \( u_k \) and \( v_k \) used in the constructive definition, respectively. The \( k \)-th singular value \( \lambda_k \) defines the \( k \)-th principal angle \( \alpha_k \) via \( \cos \alpha_k = \lambda_k \).

For \( U \) and \( V \) in \( U_{m,r} \), the chordal distance between the two subspaces \( \text{span}(U) \) and \( \text{span}(V) \) in \( G_{m,r} \) is given by \( \sqrt{\sum_{k=1}^{r} \sin^2 \alpha_k} \). The chordal distance can also be expressed in terms of singular values as \( \sqrt{\sum_{k=1}^{r} (1 - \lambda_k^2)} \). See \([13],[14]\) for more details regarding principal angles and chordal distances.

An invariant measure on \( G_{m,r} \). The space \( G_{m,r} \) admits a standard invariant measure (the Haar measure) \([15]\) in
the following sense: for any measurable set \( M \subset G \) and any \( A \in \mathcal{U}_{m,r} \), one has \( \mu(M) = \mu(AM) \), where \( A\mathcal{M} = \{ \text{span}(AU) : U \in \mathcal{U}_{m,r}, \text{span}(U) \in \mathcal{M} \} \). This invariant measure defines the uniform/isotropic distribution on the Grassmann manifold.

III. FROM THE FROBENIUS NORM TO THE GEOMETRIC METRIC

In the previous section, we showed that the matrix completion problem reduces to a search for a consistent column space. In other words, one only needs to find a global minimum of the objective function \( f_F(U) \) defined in (2). However, as we shall show in Section III-A, this approach has a serious drawback: the objective function (2) is not a continuous function of the variable \( U \). The discontinuity may prevent gradient-descent-based algorithms from converging to a global optimum [12]. To address this issue, we propose another objective function, \( f_G(U) \), based on the geometry of the problem (formally defined in Section III-B.) To solve the matrix completion problem in this setting, one then needs to solve the problem

\[
(P2): \text{find a } U \in \mathcal{U}_{m,r} \text{ such that } f_G(U) = 0. \tag{4}
\]

The new objective function \( f_G \) has two interesting properties. First, it is a continuous function of \( U \). Second, the preimage of \( f_G(U) = 0 \) is the closure of the preimage of \( f_F(U) = 0 \). This means that \( f_G \) can be viewed as the tightest “continuous approximation” of \( f_F \) (up to some scalar).

A. Why the Frobenius Norm Fails

We use an example to show that the objective function (2) based on the Frobenius norm is not continuous. Let \( x_{\Omega,i} \), be the \( i \)th column of the matrix \( X_{\Omega,i} \). Let \( \Omega_i \subset [m] \) be the set of indices of known entries in the \( i \)th column. We use \( P_{\Omega,i} \) to denote the projection operator corresponding to the index set of \( \Omega_i \). By additivity of the squared Frobenius norm, the objective function can be written as a sum of atomic functions, i.e.,

\[
f_F(U) = \sum_{i=1}^{n} \min_{w_i \in \mathbb{R}} \| x_{\Omega,i} - P_{\Omega,i}(Uw_i) \|_F^2 = \sum_{i=1}^{n} \| x_{\Omega,i} - P(x_{\Omega,i}, U_{\Omega,i}) \|_F^2,
\]

where \( U_{\Omega,i} = [P_{\Omega,i}(u_1), \ldots, P_{\Omega,i}(u_c)] \) and \( u_1, \ldots, u_c \) are column vectors of the matrix \( U \), and the symbol \( P \) in the last equality is the standard projection operator. We show in the next example that an atomic function, say \( f_{F,1}(U) \), is not continuous.

**Example 1:** Suppose that \( x_{\Omega,1} = [0, 1, 1]^T \) and \( \Omega_1 = \{2,3\} \). Let \( U \) be of the form \( U = \left[ \sqrt{1-2\nu}, \nu, \nu, \nu \right]^T \) \( \epsilon \in \mathbb{R} \), where \( \nu \in [-1/\sqrt{2}, 1/\sqrt{2}] \). For a given \( U \), the atomic function \( f_{F,1}(U) \) is given by \( f_{F,1}(U) = \min_{w_1 \in \mathbb{R}} \left\| [0, 1, 1]^T - P_{\Omega,1}(Uw_1) \right\|_F^2 \). This is a quadratic optimization problem and can be easily solved. The optimal \( w^* \) is given by

\[
w^* = \begin{cases} \frac{2}{\epsilon} & \text{if } \epsilon \neq 0, \\ 0 & \text{if } \epsilon = 0. \end{cases}
\]

Hence, one has

\[
f_{F,1}(U(\epsilon)) = \begin{cases} 0 & \text{if } \epsilon \in \left[ -\frac{1}{\sqrt{2}}, 0 \right] \cup \left[ 0, \frac{1}{\sqrt{2}} \right], \\ 2 & \text{if } \epsilon = 0. \end{cases}
\]

which shows that \( f_{F,1}(U(\epsilon)) \) has a singular point at \( \epsilon = 0 \).

It is straightforward to construct examples where the overall objective function (2) is a discontinuous function of \( U \). As has been shown in Example III-A, the solution set of \( f_F = 0 \) may not be closed.

B. A Geometric Metric

To address the problem due to the singularities of the objective functions, we propose to replace the Frobenius norm by a geometric performance metric. In this case, the objective function is defined as

\[
f_G(U) = \sum_{i=1}^{n} f_{G,i}(U),
\]

where \( f_{G,i}(U) \) denotes the geometric metric corresponding to the \( i \)th column, defined as follows. If \( x_{\Omega,i} = 0 \), we set \( f_{G,i}(U) = 0 \). Henceforth, we only consider the case when \( x_{\Omega,i} \neq 0 \). For any \( x_{\Omega,i} \neq 0 \), let \( x_{\Omega,i} = x_{\Omega,i}/\|x_{\Omega,i}\|_F \) be the normalized vector \( x_{\Omega,i} \). Let \( \Omega_i^c = \{1, 2, \ldots, m\} \setminus \Omega_i \) be the complement of \( \Omega_i \). Let \( e_k \in \mathbb{R}^m \) be the \( k \)th natural basis vector, i.e., the \( k \)th entry of \( e_k \) equals to one and all other entries are zero. Define

\[
B_i = [x_{\Omega,i}, e_{k_1}, \ldots, e_{k_t}],
\]

where \( \{k_1, \ldots, k_t\} = \Omega_i^c \). Let \( \lambda_{\text{max}}(B_i^T U) \) be the largest singular value of the matrix \( B_i^T U \). Then

\[
f_{G,i}(U) = 1 - \lambda_{\text{max}}^2(B_i^T U). \tag{6}
\]

This expression is closely related to the chordal distance between two subspaces [13, 14], we henceforth refer to the function (6) either as the geometric metric (6), or with slight abuse of terminology, as the chordal distance.

One advantage of the chordal distance is its continuity. This follows directly from the continuity of the singular values of the underlying matrix. More importantly, the following theorem shows that the preimage of \( f_{G,i}(U) = 0 \) is actually the closure of the preimage of \( f_{F,i}(U) = 0 \).

**Theorem 1:** Given \( x_{\Omega,i} \in \mathbb{R}^m \) and \( \Omega_i \subset [m] \). Let \( U_{\Omega,i} \in \mathbb{R}^{m \times r} \) be such that \( (U_{\Omega,i})_{k,\ell} = U_{k,\ell} \) if \( k \in \Omega_i \) and \( (U_{\Omega,i})_{k,\ell} = 0 \) if \( k \notin \Omega_i \). Define

\[
U_{F,i} = \left\{ U \in U_{m,r} : f_{F,i}(U) = \|x_{\Omega,i} - P(x_{\Omega,i}, U_{\Omega,i})\|_F^2 = 0 \right\}
\]

and

\[
U_{G,i} = \left\{ U \in U_{m,r} : f_{G,i}(U) = 1 - \lambda_{\text{max}}^2(B_i^T U) = 0 \right\}.
\]

Then \( U_{G,i} \) is the closure of \( U_{F,i} \), i.e., \( U_{G,i} = \overline{U_{F,i}} \).

The detailed proof can be found in the journal version of this paper [16]. Although this theorem deals with only one column of the observed matrix, the result can be easily extended to the whole matrix \( X_{\Omega} \) [16].
IV. PERFORMANCE GUARANTEES

Consider the matrix completion problem described in (4). The following theorem describes completion scenarios for which a global optimum can be found with probability one.

**Theorem 2:** Consider the following cases:

1. (rank-one matrices with arbitrary sampling): Let $X_\Omega = F_\Omega(X)$ for some unknown matrix $X$ with rank equal to one. Here, $\Omega \subset [n] \times [n]$ can be arbitrary.
2. (full sampling with arbitrary rank matrices): Let $X_\Omega = X$, i.e., $\Omega = [n] \times [n]$.

Suppose that $r = \text{rank}(X)$ is given. Let $U_G \subset U_{m,r}$ be the preimage of $f_G(U) = 0$. Let $U_0$ be randomly generated from the isotropic distribution on $U_{m,r}$, and used as the initial point of the search procedure. With probability one, there exists a continuous path $U(t)$, $t \in [0,1]$, such that $U(0) = U_0$, $U(1) \in U_G$ and $\frac{d}{dt} f_G(t) \leq 0$ for all $t \in (0,1)$, where the equality holds if and only if $U \in U_G$.

The proof of the theorem is detailed in the journal version of this paper [16]. It is based on the following two assumptions:

**Assumption I:** There exists a global optimum $X_U \in U_{m,r}$ such that $f_G(X_U) = 0$ and all the $r$ principal angles between $U_X$ and $X_U$ are less than $\pi/2$. That is, all the singular values of $U_X^T U_U$ are strictly positive.

**Assumption II:** All of the $\theta_i$’s (the smallest principal angle between span $(U_0)$ and span $(B_i)$) are less than $\pi/2$.

**Remark 1:** Suppose that the matrix $U_0$ is randomly drawn from the uniform (isotropic) distribution on $U_{m,r}$. Then $U_0$ satisfies both assumptions with probability one [15], [14].

It is worth noting that almost all starting points for gradient descent are good: if the starting point is a consistent solution, the algorithm halts immediately; otherwise, there exists a continuous path from this starting point to a global optimum such that the objective function keeps decreasing. The performance guarantee provided in Theorem 2 is strong in the sense that it does not require either incoherence conditions or large matrix sizes.

We also developed a general method to search for a consistent solution of the low-rank matrix completion problem. Under the assumptions of Theorem 2, for almost all $U_0 \in U_{m,r}$, there exists a continuous path leading to a global minimizer. However, one does not know this path in the process of solving the matrix completion problem. A practical approach is to use a gradient descent method. In the journal version of this paper [16], we present a randomized gradient descent algorithm. Let $U^{(1)} \in U_{m,r}, i = 1, 2, \ldots$, be the starting point of the $i$th iteration. Even though $U^{(1)}$’s are not independent and they are not isotropically distributed, one can show that with probability one, all $U^{(1)}$’s are “good” in the following sense: first, there is a continuous path $U(t)$ from $U^{(1)}$ to a global minimum; second, the objective function decreases after each iteration. Hence, this gradient descent procedure converges to a global minimum as the number of iterations approaches infinity. This general method has strong performance guarantees for the two cases described in Theorem 2.

A simple, yet extremely important corollary of the Theorem 2 is as follows: suppose that the partial observations $X_\Omega$ admit a unique consistent solution in terms of the Frobenius norm; then a gradient search procedure using the geometric norm finds this unique solution with probability one. This conclusion follows from the fact that the solution set under the Frobenius norm contains only a single point and therefore $U_G = U_F = U_F$.

For the more general case where $r > 1$ and $\Omega \neq [n] \times [n]$, we cannot guarantee the same performance guarantees. Nevertheless, in the following corollary, we present a collection of results that may be helpful for future exploration.

**Corollary 1:** (General Cases) Let $X \in \mathbb{R}^{m \times n}$ be a rank-$r$ matrix. Let $U_X \in U_G$ be a global minimum. For each $i \in [n]$, the following statements are true. Let $U_{t,0} \in U_{m,r}$ and $w_i \in U_{r,1}$ be randomly drawn from the corresponding isotropic distributions respectively. Then with probability one, the vector $U_{0,i} \pm U_{t,0} w_i$ is not orthogonal to $U_{X,i}$. Suppose that this is true. Define $\theta_i = \cos^{-1} |\langle U_{t}(0), B_i \rangle|$. There exists a continuous path $U(t)$ in $U_{m,1}$ such that $U(t)(0) = U_{0,i}$, $U(t)(1) \in \text{span}(U_{X,t}) \cap U_{m,1}$, and $\frac{d}{dt} \sin^2 \theta_i \leq 0$, where the equality holds if and only if $\theta_i(t) = 0$.

**Remark 2:** This corollary is similar to Theorem 2 in the sense that there exist continuous paths along which the atomic functions decreases. However, the paths $U(t)$ in Corollary 1 may be different for different $i$’s, while in Theorem 2, a single continuous path $U(t)$ is constructed.

REFERENCES