A New Retraction for Accelerating the Riemannian Three-Factor Low-Rank Matrix Completion Algorithm

Zhizhong Li[1,2], Deli Zhao[2], Zhouchen Lin[3,4]*, Edward Y. Chang[5]
[1]School of Mathematics, Peking University, Beijing, China
[2]Advanced Algorithm Research Group, HTC, Beijing, China
[3]Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, Beijing, China
[4]Cooperative Medianet Innovation Center, Shanghai, China
[5]HTC Research
lizz@pku.edu.cn, zhaodeli@gmail.com, zlin@pku.edu.cn, eyuchang@gmail.com

Abstract

The Riemannian three-factor matrix completion (R3MC) algorithm is one of the state-of-the-art geometric optimization methods for the low-rank matrix completion problem. It is a nonlinear conjugate-gradient method optimizing on a quotient Riemannian manifold. In the line search step, R3MC approximates the minimum point on the searching curve by minimizing on the line tangent to the curve. However, finding the exact minimum point by iteration is too expensive. We address this issue by proposing a new retraction with a minimizing property. This special property provides the exact minimization for the line search by establishing correspondences between points on the searching curve and points on the tangent line. Accelerated R3MC, which is R3MC equipped with this new retraction, outperforms the original algorithm and other geometric algorithms for matrix completion in our empirical study.

1. Introduction

The problem of low-rank matrix completion is completing a matrix $X^* \in \mathbb{R}^{n\times m}$ from a small number of known entries under the low-rank assumption. Applications of matrix completion pervade in machine learning [9], computer vision [10], system identification [5], and so on. One formulation of the matrix completion problem [3] consists of finding the lowest rank matrix that agrees with known entries of X^*. The other formulation takes the presence of noise into consideration and admits misfit within a given tolerance ϵ. We focus on the second approach with the rank information given, *i.e.*, finding a rank-r matrix X such that

$$\min_{X\in\mathbb{R}_r^{n\times m}} f(X) := \frac{1}{|\Omega|}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(X^*)\|_F^2, \quad (1)$$

where $\mathbb{R}_r^{n\times m}$ is the set of rank-r matrices of size $n \times m$, $\Omega := \{(i,j)|X^*_{i,j} \text{ is given}\}$ is the set of indices of known entries and $|\Omega|$ the cardinality of Ω. $\|A\|_F$ is the Frobenius norm of matrix A, and $\mathcal{P}_\Omega$ is the orthogonal sampling operator defined as

$$\mathcal{P}_\Omega(X)_{i,j} = \begin{cases} X_{i,j} & \text{when } (i,j)\in\Omega, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

We assume that the solution to Equation (1) exists.

It is well known in differential geometry that the set of rank-r $n \times m$ matrices $\mathbb{R}_r^{n\times m}$ forms a submanifold of the nm-dimensional vector space $\mathbb{R}^{nm\times 1}$. The fixed-rank matrix completion problem can be cast as a geometric optimization problem because the rank constraint could be incorporated into the manifold structure. In this point of view, the problem is turned into an unconstrained optimization. Notice that the dimension of $\mathbb{R}_r^{n\times m}$, namely $(n+m-r)r$, is much smaller than that of the ambient space $\mathbb{R}^{n\times m}$ in the case $r \ll \min(n,m)$, thus making geometric methods suitable to solve large-scale matrix completion problems.

There are a few geometric optimization algorithms available for the matrix completion problem such as LRGeomCG [11], RTRMC [2], ScGrass-CG [8], and R3MC [7], *etc*. They differ in manifold structures, Riemannian metrics and optimization methods.

LRGeomCG is a nonlinear conjugate-gradient method over the embedded submanifold $\mathbb{R}_r^{n\times m} \subset \mathbb{R}^{n\times m}$, which inherits the canonical Riemannian metric of $\mathbb{R}^{n\times m}$. The algorithm takes the embedding point of view of $\mathbb{R}_r^{n\times m}$.

*Corresponding author.

RTRMC exploits the factorization $X = UV^t$, where $U \in \mathbb{R}^{n\times r}$ and $V \in \mathbb{R}^{m\times r}$, and U is chosen to be orthonormal. It recasts the problem on the Grassmann manifold $\mathcal{G}(r,n)$ and solves the optimization using Riemannian trust-region method.

ScGrass-CG constructs the searching manifold as the product of two Grassmann manifolds, $\mathcal{G}(r,n) \times \mathcal{G}(r,m)$. The authors of ScGrass-CG also propose a scaled metric and harness a conjugate-gradient method to minimize the cost function.

R3MC, the Riemannian three-factor algorithm for low-rank matrix completion, is a geometric optimization technique using the nonlinear conjugate-gradient method on a quotient Riemannian manifold. It utilizes the factorization $X = URV^T$ to construct the searching manifold $\mathcal{M}$. The quotient manifold structure of $\mathcal{M}$ captures the symmetry of the factorization and the metric of this manifold further takes advantage of Hessian information of the cost function. R3MC is both robust and efficient, and it outperforms the state-of-the-art algorithms for addressing the matrix completion problem.

Motivated from both the geometric and the optimization perspectives, we improve R3MC by finding the exact minima in the line-search subroutine of the conjugate-gradient method. Instead of approximating the minima by inner iterations, we achieve this goal by proposing a new retraction, which is more suitable for the task of minimization as described in Proposition 3. Our numerical experiments show that the design of a suitable retraction, which helps the line search, could boost the performance of an existing geometric optimization method.

The rest of the paper is organized as follows. We first review the basic geometric notions in Section 2. Then in Section 3, we describe the implementation of the original R3MC algorithm. We present our improvements in Section 4. We show numerical experiments Section 5. In Section 6, we give our concluding remarks.

2. Geometric Optimization Framework

Optimization on manifolds is a natural generalization of traditional optimization methods on Euclidean spaces. In this section, we introduce the geometric framework for optimization on a quotient manifold and illustrate the basic geometric ingredients using R3MC as one of the examples. A detailed elaboration of optimization algorithms on matrix manifolds can be found in [1]. Readers could refer to [4] for mathematical definitions of notions in Riemannian geometry and [7] for implementation details of R3MC.

2.1. Manifold Structure

R3MC performs on the manifold

$$\mathcal{M} := (\mathrm{St}(r,n) \times \mathrm{GL}(r) \times \mathrm{St}(r,m))/(\mathcal{O}(r) \times \mathcal{O}(r)), \quad (3)$$

where $\mathrm{St}(r,n) := \{X \in \mathbb{R}^{n\times r} | X^T X = I_r\}$ is a Stiefel manifold, $\mathrm{GL}(r) := \{X \in \mathbb{R}^{r\times r} || X| \neq 0\}$ is a general linear group and $\mathcal{O}(r) := \{X \in \mathbb{R}^{r\times r} | X^T X = I_r\}$ is an orthogonal group. $|X|$ is the determinant of matrix X and I_r is the identity matrix of size $r \times r$. Let $(O_1, O_2) \in \mathcal{O}(r) \times \mathcal{O}(r)$ be an element of Lie group acting on the product manifold $\overline{\mathcal{M}} := \mathrm{St}(r,n) \times \mathrm{GL}(r) \times \mathrm{St}(r,m)$ as

$$\begin{aligned} &(\mathcal{O}(r) \times \mathcal{O}(r)) \times \overline{\mathcal{M}} \to \overline{\mathcal{M}} \\ &((O_1, O_2), (U, R, V)) \mapsto (UO_1, O_1^T R O_2, VO_2), \end{aligned} \quad (4)$$

where $(U, R, V) \in \overline{\mathcal{M}}$. We denote the canonical projection map of $\overline{\mathcal{M}}$ quotient by the group action as

$$\pi : \overline{\mathcal{M}} \to \mathcal{M}, \quad (5)$$

where $\mathcal{M}$ is the quotient space and $\overline{\mathcal{M}}$ the *total space*.

This quotient manifold structure originates from the three factor model of rank-r matrices

$$X = URV^T, \quad (6)$$

where $X \in \mathbb{R}_r^{n\times m}$ is of rank r and $(U, R, V) \in \overline{\mathcal{M}}$. The factorization is non-unique and is invariant under the group action (4). In particular, there is a homeomorphism between the following two manifolds

$$\mathbb{R}_r^{n\times m} \simeq \mathcal{M} = \overline{\mathcal{M}}/(\mathcal{O}(r) \times \mathcal{O}(r)). \quad (7)$$

It is convenient to represent an element $x \in \mathcal{M}$ of the abstract quotient manifold $\mathcal{M}$ by an element

$$\bar{x} = (U, R, V) \in \pi^{-1}(x) \subset \overline{\mathcal{M}} \quad (8)$$

in the total space $\overline{\mathcal{M}}$, which has a matrix form. Correspondingly, the tangent spaces for points x and $\bar{x}$, say $T_x\mathcal{M}$ and $T_{\bar{x}}\overline{\mathcal{M}}$, can have certain relations established through the *horizontal space*, a concept we will address in Section 2.3.

Using identification (7) and the canonical projection (5), cost function (1) can be lifted to the total space $\overline{\mathcal{M}}$ as

$$\begin{aligned} \bar{f}(U, R, V) &:= f(URV^T) \\ &= \frac{1}{|\Omega|} \|\mathcal{P}_\Omega(URV^T) - \mathcal{P}_\Omega(X^*)\|_F^2. \end{aligned} \quad (9)$$

2.2. Riemannian Metric

The Riemannian metric g on $\mathcal{M}$ is induced by the metric $\bar{g}$ [7, Section II] on the total space $\overline{\mathcal{M}}$. The definition is

$$\begin{aligned} \bar{g}_{\bar{x}}(\bar{\xi}_{\bar{x}}, \bar{\eta}_{\bar{x}}) := \mathrm{tr}((RR^T)\bar{\xi}_U^T \bar{\eta}_U) + \mathrm{tr}(\bar{\xi}_R^T \bar{\eta}_R) \\ + \mathrm{tr}((R^T R)\bar{\xi}_V^T \bar{\eta}_V), \end{aligned} \quad (10)$$

where $\bar{x} = (U, R, V) \in \overline{\mathcal{M}}$, $\bar{\xi}_{\bar{x}} = (\bar{\xi}_U, \bar{\xi}_R, \bar{\xi}_V) \in T_{\bar{x}}\overline{\mathcal{M}}$ and $\bar{\eta}_{\bar{x}} \in T_{\bar{x}}\overline{\mathcal{M}}$. This metric captures the Hessian information of the expression $\|URV^T - X^*\|_F^2$ and serves as an adaptive preconditioner for optimization.

2.3. Horizontal Space

Horizontal space $\mathcal{H}_{\bar{x}}$ at $\bar{x} \in \overline{\mathcal{M}}$ is the orthogonal complement of the vertical space $\mathcal{V}_{\bar{x}} := T_{\bar{x}}\pi^{-1}(x)$ in $T_{\bar{x}}\overline{\mathcal{M}}$.

Tangent space $T_x\mathcal{M}$ at point x in the quotient space $\mathcal{M}$ is identified with the horizontal space $\mathcal{H}_{\bar{x}} \subset T_{\bar{x}}\overline{\mathcal{M}}$ of point $\bar{x} \in \pi^{-1}(x)$ in the total space $\overline{\mathcal{M}}$ through the *horizontal lift* [1, Section 3.6.2]. The projection [7, Section III] of a tangent space $T_{\bar{x}}\overline{\mathcal{M}}$ onto the horizontal space $\mathcal{H}_{\bar{x}}$ is denoted by the operator

$$\Pi_{\bar{x}} : T_{\bar{x}}\overline{\mathcal{M}} \to \mathcal{H}_{\bar{x}}. \tag{11}$$

2.4. Ambient Space

Total space $\overline{\mathcal{M}}$ is a submanifold of the ambient vector space

$$\mathcal{E} := \mathbb{R}^{n\times r} \times \mathbb{R}^{r\times r} \times \mathbb{R}^{m\times r}. \tag{12}$$

Intermediate steps of the optimization algorithms, for example, computing vector transports as shown in Equation (20) may produce vectors in $T_{\bar{x}}\mathcal{E}$ that do not belong to $T_{\bar{x}}\overline{\mathcal{M}}$. The linear operator Ψ [7, Section III],

$$\Psi_{\bar{x}} : T_{\bar{x}}\mathcal{E} \to T_{\bar{x}}\overline{\mathcal{M}}, \tag{13}$$

is used to project vectors in $T_{\bar{x}}\mathcal{E}$ orthogonally to the subspace $T_{\bar{x}}\overline{\mathcal{M}}$.

2.5. Retraction

Retraction [1, Definition 4.1.1] is an analog of the geometric exponential map providing a way to move along a direction while constraint to the manifold. A retraction $\widetilde{\mathcal{R}}$ should satisfy two conditions: 1) Centering: $\widetilde{\mathcal{R}}_{\bar{x}}(0_{\bar{x}}) = \bar{x}$, where the vector $0_{\bar{x}}$ is the zero element of $T_{\bar{x}}\overline{\mathcal{M}}$; and 2) Local rigidity:

$$\mathrm{D}\widetilde{\mathcal{R}}_{\bar{x}}(0_{\bar{x}}) = \mathrm{id}_{T_{\bar{x}}\overline{\mathcal{M}}}, \tag{14}$$

where $\mathrm{D}\widetilde{\mathcal{R}}_{\bar{x}}(0_{\bar{x}})$ is the differential of the map $\widetilde{\mathcal{R}}_{\bar{x}}$ at the zero element $0_{\bar{x}} \in T_{0_{\bar{x}}}T_{\bar{x}}\overline{\mathcal{M}} \simeq T_{\bar{x}}\overline{\mathcal{M}}$. In R3MC, the retraction $\overline{\mathcal{R}}$ at point $\bar{x} = (U, R, V) \in \overline{\mathcal{M}}$ is a map

$$\begin{aligned} \overline{\mathcal{R}}_{\bar{x}} : T_{\bar{x}}\overline{\mathcal{M}} &\to \overline{\mathcal{M}} \\ \bar{\xi}_{\bar{x}} &\mapsto (\mathrm{uf}(U + \bar{\xi}_U), R + \bar{\xi}_R, \mathrm{uf}(V + \bar{\xi}_V)), \end{aligned} \tag{15}$$

where $\bar{\xi}_{\bar{x}} = (\bar{\xi}_U, \bar{\xi}_R, \bar{\xi}_V) \in T_{\bar{x}}\overline{\mathcal{M}}$ is a given direction and

$$\mathrm{uf}(A) := A(A^TA)^{-1/2} \tag{16}$$

is the orthogonal factor of the full column-rank matrix A in polar decomposition.

To get a sense how it works, we illustrate the behavior of $\overline{\mathcal{R}}$ in a low-dimensional toy model.

Example 1. Consider a simple manifold

$$\overline{\mathcal{M}}_0 := \mathrm{St}(1,2) \times \mathrm{GL}(1) \times \mathrm{St}(1,1) \simeq S^1 \times \mathbb{R}_* \times \{\pm 1\}, \tag{17}$$

where $\mathrm{St}(1,2) := \{X \in \mathbb{R}^2 | X^TX = 1\} \simeq S^1$, $\mathbb{R}_* := \{x \in \mathbb{R} | x \neq 0\}$ and $\mathrm{St}(1,1) = \{\pm 1\}$. Pick one component

$$\overline{\mathcal{M}}_0^+ := S^1 \times \mathbb{R}_+ \times \{1\} \simeq S^1 \times \mathbb{R}_+ \subset \mathbb{R}^3, \tag{18}$$

where $\mathbb{R}_+ := \{x \in \mathbb{R} | x > 0\}$ and $\overline{\mathcal{M}}_0^+$ is identified with a cylinder in $\mathbb{R}^3$. Using coordinates of $\mathbb{R}^3$, retraction $\overline{\mathcal{R}}$ at $\bar{x} = (x_0, y_0, z_0)$ is

$$\overline{\mathcal{R}}_{\bar{x}}(\bar{\xi}_{\bar{x}}) = ((x_0 + \xi_1)/l, (y_0 + \xi_2)/l, z_0 + \xi_3), \tag{19}$$

where $\bar{\xi}_{\bar{x}} = (\xi_1, \xi_2, \xi_3)$ and $l := ((x_0 + \xi_1)^2 + (y_0 + \xi_2)^2)^{1/2}$. In the equation, $\overline{\mathcal{R}}$ keeps z coordinate of $\bar{x} + \bar{\xi}_{\bar{x}}$ fixed and scales the other two coordinates to the unit circle parallel to the x-y plane, as shown in Figure 1.

Retracting the line $t \mapsto t\bar{\xi}_{\bar{x}}$ in $T_{\bar{x}}\overline{\mathcal{M}}$, we'll get a curve $\phi(t) := \overline{\mathcal{R}}_{\bar{x}}(t\bar{\xi}_{\bar{x}})$ on $\overline{\mathcal{M}}_0^+$ which plays the role of a geodesic. Tangent line $\gamma(t) := \bar{x} + t\bar{\xi}_{\bar{x}}$ is the first-order approximation of the curve $\phi(t)$. Moving along direction $\bar{\xi}_{\bar{x}}$ means moving along the curve $\phi(t)$ and parameter t is the *step-size*. □

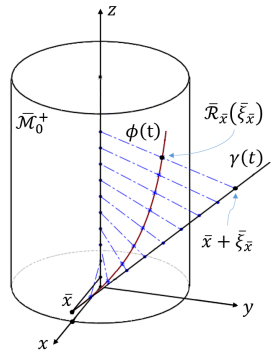

Figure 1. Illustration of Example 1: Retraction $\overline{\mathcal{R}}$ on the cylinder $\overline{\mathcal{M}}_0^+$. The definitions of $\overline{\mathcal{M}}_0^+$ and $\overline{\mathcal{R}}$ are in Equations (18) and (19), respectively. $\bar{\xi}_{\bar{x}}$ is a direction. $\bar{x}$ and $\overline{\mathcal{R}}_{\bar{x}}(\bar{\xi}_{\bar{x}})$ are points on the curve $\phi(t)$, which both lie on the cylinder. $\gamma(t)$ is a straight line tangent to $\phi(t)$ at point $\bar{x}$.

2.6. Vector Transport

Vector transport [1, Definition 8.1.1] is an analog of parallel transport in geometry. It moves a vector from the tangent space of one point to the tangent space of another one. In R3MC, vector transport $\mathcal{T}_{\eta_x}\xi_x$ on manifold $\mathcal{M}$ could be written as

$$\overline{\mathcal{T}_{\eta_x}\xi_x} = \Pi_{\bar{y}}(\Psi_{\bar{y}}(\bar{\xi}_{\bar{x}})), \tag{20}$$

where $\bar{x} \in \pi^{-1}(x)$ and $\bar{y} := \overline{\mathcal{R}}_{\bar{x}}(\bar{\eta}_{\bar{x}})$. $\bar{\eta}_{\bar{x}}$, $\bar{\xi}_{\bar{x}}$ and $\overline{\mathcal{T}_{\eta_x}\xi_x}$ are horizontal lifts of η_x, ξ_x, and $\mathcal{T}_{\eta_x}\xi_x$, respectively. In the expression $\Psi_{\bar{y}}(\bar{\xi}_{\bar{x}})$, vector $\bar{\xi}_{\bar{x}}$ is seen as an element of the tangent space $T_{\bar{y}}\mathcal{E}$. Equation (20) shows that the transport $\mathcal{T}_{\eta_x}\xi_x$ of vector $\xi_x \in T_x\mathcal{M}$ from point x to point $y := \pi(\bar{y})$ can be obtained by translating the lifted vector $\bar{\xi}_{\bar{x}} \in \mathcal{H}_{\bar{x}} \subset T_{\bar{x}}\mathcal{E}$ from point $\bar{x}$ to point $\bar{y}$, and then projecting the new $\bar{\xi}_{\bar{x}} \in T_{\bar{y}}\mathcal{E}$ by $\Psi_{\bar{y}}$ and $\Pi_{\bar{y}}$ to the horizontal space $\mathcal{H}_{\bar{y}}$.

3. Optimization Algorithm of R3MC

Theoretically, R3MC is running on the abstract quotient manifold $\mathcal{M}$. In order to do computation in simple matrix forms, it is convenient to employ the corresponding lifted objects (with 'bar' notations) on $\overline{\mathcal{M}}$ as working variables.

The original R3MC method [7, Algorithm 1] is shown in Algorithm 1.

Algorithm 1 The Original R3MC Method

Input: Given an initial point $\bar{x}_0 \in \overline{\mathcal{M}}$.

1: Compute Riemannian gradient $\bar{\xi}_i = \overline{\text{grad}}_{\bar{x}_i} \bar{f} \in \mathcal{H}_{\bar{x}_i}$.
2: Compute the conjugate direction $\bar{\eta}_i \in \mathcal{H}_{\bar{x}_i} \subset T_{\bar{x}_i}\overline{\mathcal{M}}$ by Polak-Ribière+ .
3: If $\bar{\eta}_i$ is not a descent direction, reset $\bar{\eta}_i = -\bar{\xi}_i$.
4: Compute an initial step-size s_i.
5: Perform the Armijo backtracking and modify s_i when necessary.
6: Retract along $\bar{\eta}_i$ with step-size s_i to iterate $\bar{x}_{i+1}$.
7: Repeat until convergence.

In the above algorithm, Riemannian gradient $\bar{\xi}_i$ is the lift of the gradient of f on $\mathcal{M}$ to the total space $\overline{\mathcal{M}}$. Conjugate direction $\bar{\eta}_i$ is defined as

$$\bar{\eta}_i = -\bar{\xi}_i + \beta_i \Pi_{\bar{x}_i}(\Psi_{\bar{x}_i}(\bar{\eta}_{i-1})), \tag{21}$$

where $\Pi_{\bar{x}_i}(\Psi_{\bar{x}_i}(\bar{\eta}_{i-1}))$ is the transport of vector $\bar{\eta}_{i-1}$ from point $\bar{x}_{i-1}$ to $\bar{x}_i$ as in Equation (20) and β_i is computed by the Polak-Ribière+ method

$$\beta_i = \frac{\langle \bar{\xi}_i, \bar{\xi}_i - \mathcal{T}_{s_{i-1}\bar{\eta}_{i-1}}(\bar{\xi}_{i-1})\rangle}{\langle \bar{\xi}_{i-1}, \bar{\xi}_{i-1}\rangle}. \tag{22}$$

The initial step-size s_i is estimated by minimizing a degree 6 polynomial $\bar{f}(t) := \bar{f}(\bar{x}_i + t\bar{\eta}_i)$ in variable t which is the cost function $\bar{f}$ restricted on the tangent line $\gamma(t) = \bar{x}_i + t\bar{\eta}_i$. R3MC can apply the degree 2 approximation of $\bar{f}(t)$ which is called the *accelerated linearized step-size guess*.

This optimization process is illustrated on the toy model $\overline{\mathcal{M}}_0^+$, as shown in Figure 2.

We will define a new retraction $\widetilde{\mathcal{R}}$ on $\overline{\mathcal{M}}$ with a good minimizing property in the next section. This special property ensures that minimization on the retracted curve is equivalent to minimization along the tangent line. So the exact minimum point could be found and this leads to a robust performance for the completion of ill-conditioned matrices.

4. Accelerated R3MC

We accelerate R3MC by changing the line-search strategy from a rough initial guess to an exact minimization. This is done by introducing a new retraction with good property of minimization such that we can find the exact minima without computation overhead.

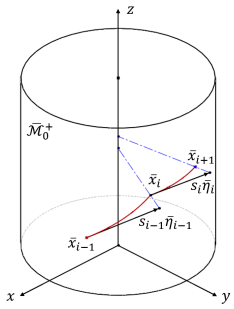

Figure 2. R3MC optimization on the manifold $\overline{\mathcal{M}}_0^+$. Starting from point $\bar{x}_{i-1}$, two consecutive steps with search directions $\bar{\eta}_{i-1}$ and $\bar{\eta}_i$ are drawn. The algorithm goes to point $\bar{x}_i$ and $\bar{x}_{i+1}$ using s_{i-1} and s_i as step-sizes, respectively. Point $s_{i-1}\bar{\eta}_{i-1}$ is the minimum point of $\bar{f}$ along the tangent line through $\bar{x}_{i-1}$. However, the retracted point $\bar{x}_i$ is just an approximation to the minimum point along the retracted curve on the cylinder.

There are two motivations for tuning the line-search step: 1) As suggested in [1, Section 8.3], exact minimum value should always be employed if computation overhead is not prohibitive because the next search direction would more likely be a decent direction. 2) The selection of step-size has a great influence on geometric optimization algorithms. As observed in [11, Section 3], a good initial guess assures that the backtracking step is almost unnecessary.

4.1. A-R3MC with New Retraction

Our new retraction $\widetilde{\mathcal{R}}$ at point $\bar{x} = (U, R, V)$ on $\overline{\mathcal{M}}$ is defined as

$$\begin{aligned} \widetilde{\mathcal{R}}_{\bar{x}} : T_{\bar{x}}\overline{\mathcal{M}} &\to \overline{\mathcal{M}} \\ \bar{\xi}_{\bar{x}} &\mapsto (P_1, Q_1(R + \bar{\xi}_R)Q_2^T, P_2), \end{aligned} \tag{23}$$

where $\bar{\xi}_{\bar{x}} = (\bar{\xi}_U, \bar{\xi}_R, \bar{\xi}_V)$ is the given direction and P_1, Q_1, P_2 and Q_2 are defined by polar decompositions $U + \bar{\xi}_U = P_1Q_1$ and $V + \bar{\xi}_V = P_2Q_2$, respectively. In fact

$$\begin{aligned} P_1 &= \text{uf}(U + \bar{\xi}_U) \\ &= (U + \bar{\xi}_U)((U + \bar{\xi}_U)^T(U + \bar{\xi}_U))^{-1/2} \\ &= (U + \bar{\xi}_U)(I_r + U^T\bar{\xi}_U + \bar{\xi}_U^T U + \bar{\xi}_U^T\bar{\xi}_U)^{-1/2} \\ &= (U + \bar{\xi}_U)(I_r + \bar{\xi}_U^T\bar{\xi}_U)^{-1/2} \end{aligned} \tag{24}$$

and $Q_1 = (I_r + \bar{\xi}_U^T\bar{\xi}_U)^{1/2}$, where we use the characterization of the tangent space of a Stiefel manifold [1, Example 3.5.2]: $U^T\bar{\xi}_U + \bar{\xi}_U^T U = 0$. Similarly, $P_2 = \text{uf}(V + \bar{\xi}_V) = (V + \bar{\xi}_V)(I_r + \bar{\xi}_V^T\bar{\xi}_V)$.

First we prove that the expression (23) indeed defines a retraction on $\overline{\mathcal{M}}$.

Proposition 1. *$\widetilde{\mathcal{R}}$ is a retraction on $\overline{\mathcal{M}}$.*

Proof. The first condition, $\widetilde{\mathcal{R}}_{\bar{x}}(0_{\bar{x}}) = \bar{x}$, for $\widetilde{\mathcal{R}}$ being a retraction is obvious. Now we verify the local rigidity condition (14). Let $\bar{x} = (U, R, V)$, $\bar{\xi}_{\bar{x}} = (\bar{\xi}_U, \bar{\xi}_R, \bar{\xi}_V) \in T_{\bar{x}}\overline{\mathcal{M}}$. Define curve $\gamma(t) := t\bar{\xi}_{\bar{x}}$ in $T_{\bar{x}}\overline{\mathcal{M}}$, then $\gamma(0) = 0_{\bar{x}}$ and $\dot{\gamma}(0) = \bar{\xi}_{\bar{x}}$. Denote $\widetilde{\mathcal{R}}_{\bar{x}}(\gamma(t)) =: (A(t), B(t), C(t))$ where

$$\begin{aligned} A(t) &= (U + t\bar{\xi}_U)(I_r + t^2\bar{\xi}_U^T\bar{\xi}_U)^{-1/2} \\ &= U + t\bar{\xi}_U - \frac{t^2}{2}U\bar{\xi}_U^T\bar{\xi}_U + \cdots, \qquad (25a) \\ B(t) &= (I_r + t^2\bar{\xi}_U^T\bar{\xi}_U)^{1/2}(R + t\bar{\xi}_R)(I_r + t^2\bar{\xi}_V^T\bar{\xi}_V)^{1/2} \\ &= R + t\bar{\xi}_R + \frac{t^2}{2}(R\bar{\xi}_V^T\bar{\xi}_V + \bar{\xi}_U^T\bar{\xi}_U R) + \cdots, \qquad (25b) \\ C(t) &= (V + t\bar{\xi}_V)(I_r + t^2\bar{\xi}_V^T\bar{\xi}_V)^{-1/2} \\ &= V + t\bar{\xi}_V - \frac{t^2}{2}V\bar{\xi}_V^T\bar{\xi}_V + \cdots. \qquad (25c) \end{aligned}$$

The first line of (25a) uses Equation (24) and the second line uses Taylor expansion. So we have

$$\begin{aligned} \mathrm{D}\widetilde{\mathcal{R}}_{\bar{x}}(0_{\bar{x}})(\bar{\xi}_{\bar{x}}) &= \left.\frac{d}{dt}\right|_{t=0} \widetilde{\mathcal{R}}_{\bar{x}}(\gamma(t)) \\ &= (\dot{A}(0), \dot{B}(0), \dot{B}(0)) = \bar{\xi}_{\bar{x}}. \end{aligned} \qquad (26)$$

This means that condition (14) is satisfied. □

We illustrate the new retraction together with the original one on the example manifold $\overline{\mathcal{M}}_0^+$ in Figure 3. Two curves drawn on the cylinder show the different behaviors of these retractions. Notice that points on the blue dashed curve $\psi(t)$ connecting $\gamma(t)$ and $\phi(t)$ share the same function value of $\bar{f}$. In fact, suppose $\bar{x} = (x_0, y_0, z_0)$ and $\bar{\xi}_{\bar{x}} = (\xi_1, \xi_2, \xi_3) \in T_{\bar{x}}\overline{\mathcal{M}}_0^+$, then the curve $\psi(t)$ passing through points $\bar{x} + \bar{\xi}_{\bar{x}}$ and $\widetilde{\mathcal{R}}_{\bar{x}}(\bar{\xi}_{\bar{x}})$ has the expression

$$\psi(t) : t \mapsto ((x_0 + \xi_1)/t, (y_0 + \xi_2)/t, t(z_0 + \xi_3)), \qquad (27)$$

where t ranges from 1 to $((x_0 + \xi_1)^2 + (y_0 + \xi_2)^2)^{1/2}$. So the identity $\bar{f}(\psi(t)) \equiv \bar{f}(x_0 + \xi_1, y_0 + \xi_2)$ is apparent by simple computations. This observation is the motivation for Propostion 3 later in this section.

Since the algorithm performs on the quotient manifold $\mathcal{M}$, the new retraction $\widetilde{\mathcal{R}}$ should be compatible with the quotient structure of $\overline{\mathcal{M}}$, in the sense that it can induce a retraction $\mathcal{R}$ on $\mathcal{M}$,

$$\begin{aligned} \mathcal{R}_x : T_x\mathcal{M} &\to \mathcal{M} \\ \xi_x &\mapsto \pi(\widetilde{\mathcal{R}}_{\bar{x}}(\bar{\xi}_{\bar{x}})), \end{aligned} \qquad (28)$$

where $\bar{x} \in \pi^{-1}(x)$ and $\bar{\xi}_{\bar{x}}$ is the horizontal lift of ξ_x.

Proposition 2. *$\mathcal{R}$ is a well-defined retraction on $\mathcal{M}$.*

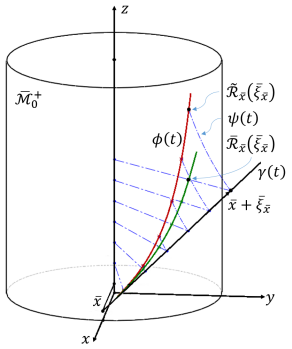

Figure 3. New retraction $\widetilde{\mathcal{R}}$ and the original retraction $\overline{\mathcal{R}}$.In this figure, $\bar{x}$ is a point on the cylinder $\overline{\mathcal{M}}_0^+$, $\gamma(t)$ a line passing through $\bar{x}$ with direction $\bar{\xi}_{\bar{x}}$. The upper red curve $\phi(t)$ is the retracted curve $\phi(t) := \widetilde{\mathcal{R}}_{\bar{x}}(t\bar{\xi}_{\bar{x}})$ under the new retraction $\widetilde{\mathcal{R}}$, and the lower green curve is the one generated by the retraction $\overline{\mathcal{R}}$. The curve $\psi(t)$ is defined in Equation (27).

Proof. We verify that $\widetilde{\mathcal{R}}$ satisfies the condition of Proposition 4.1.3 in [1] which reads

$$\pi(\widetilde{\mathcal{R}}_{\bar{x}_a}(\bar{\xi}_{\bar{x}_a})) = \pi(\widetilde{\mathcal{R}}_{\bar{x}_b}(\bar{\xi}_{\bar{x}_b})), \qquad (29)$$

where $\bar{x}_a$ and $\bar{x}_b$ are any two points in the fiber $\pi^{-1}(x)$ for point $x \in \mathcal{M}$, and $\bar{\xi}_{\bar{x}_a} \in \mathcal{H}_{\bar{x}_a}$ and $\bar{\xi}_{\bar{x}_b} \in \mathcal{H}_{\bar{x}_b}$ are both horizontal lifts of $\xi_x \in T_x\mathcal{M}$. Let $\bar{x}_a = (U_a, R_a, V_a)$ and $\bar{\xi}_{\bar{x}_a} = (\bar{\xi}_{U_a}, \bar{\xi}_{R_a}, \bar{\xi}_{V_a})$, then $\bar{x}_b$ can be written as $(U_aO_1, O_1^TR_aO_2, V_aO_2)$ for some orthogonal matrices $O_1, O_2 \in \mathcal{O}(r)$. Now compute the expression for $\bar{\xi}_{\bar{x}_b}$. Suppose $\gamma_a(t) = (A(t), B(t), C(t))$ is a curve in $\overline{\mathcal{M}}$ such that $\gamma_a(0) = \bar{x}_a$ and $\dot{\gamma}_a(0) = \bar{\xi}_{\bar{x}_a}$. Define curve

$$\gamma_b(t) := (A(t)O_1, O_1^TB(t)O_2, C(t)O_2), \qquad (30)$$

then $\gamma_b(0) = \bar{x}_b$, and

$$\dot{\gamma}_b(0) = (\bar{\xi}_{U_a}O_1, O_1^T\bar{\xi}_{R_a}O_2, \bar{\xi}_{V_a}O_2). \qquad (31)$$

It is easy to check that the tangent vector $\dot{\gamma}_b(0)$ fulfills the requirements for the horizontal space $\mathcal{H}_{\bar{x}_b}$ [7, Section III]. Since $\pi(\gamma_a(t)) = \pi(\gamma_b(t))$,

$$\mathrm{D}\pi(\bar{x}_b)(\dot{\gamma}_b(0)) = \mathrm{D}\pi(\bar{x}_a)(\dot{\gamma}_a(0)) = \xi_x, \qquad (32)$$

where $\mathrm{D}\pi(\bar{x}_b)(\dot{\gamma}_b(0))$ is the image of tangent vector $\dot{\gamma}_b(0)$ under the differential of canonical projection π at point $\bar{x}_b$. The second equal sign comes from the fact that $\dot{\gamma}_a(0) = \bar{\xi}_{\bar{x}_a}$ is the horizontal lift of ξ_x. So Equation (32) shows that $\dot{\gamma}_b(0) \in \mathcal{H}_{\bar{x}_b}$ is the horizontal lift of ξ_x at $\bar{x}_b$ and we have $\bar{\xi}_{\bar{x}_b} = \dot{\gamma}_b(0)$.

If $U_a + \bar{\xi}_{U_a} = P_1Q_1$ and $V_a + \bar{\xi}_{V_a} = P_2Q_2$ by polar decomposition, then $U_aO_1 + \bar{\xi}_{U_a}O_1 = (P_1O_1)(O_1^TQ_1O_1)$, and $V_aO_2 + \bar{\xi}_{V_a}O_2 = (P_2O_2)(O_2^TQ_2O_2)$. By Equation (23), $\widetilde{\mathcal{R}}_{\bar{x}_a}(\bar{\xi}_{\bar{x}_a}) = (P_1, Q_1(R_a + \bar{\xi}_{R_a})Q_2^T, P_2)$ and

$$\begin{aligned} \widetilde{\mathcal{R}}_{\bar{x}_b}(\bar{\xi}_{\bar{x}_b}) = (P_1O_1, O_1^TQ_1O_1(O_1^TR_aO_2 \\ + O_1^T\bar{\xi}_{R_a}O_2)O_2^TQ_2^TO_2, P_2O_2). \end{aligned} \qquad (33)$$

Now it is easy to check that Equation (29) holds. □

We come back to Figure 3. Recall that the invariance of the cost function on curve ψ establishes a correspondence between points on γ and points on ϕ. This correspondence in fact can induce an minimizing property that allow exact minimizations for the line-search steps of Algorithm 1. This minimizing property is stated as follows:

Proposition 3. *Suppose that $\bar{x} \in \overline{\mathcal{M}}$ and $\bar{\eta}_{\bar{x}} \in \mathcal{H}_{\bar{x}}$. Let $\phi(t) := \widetilde{\mathcal{R}}_{\bar{x}}(t\bar{\eta}_{\bar{x}})$ be the retracted curve of line $t \mapsto t\bar{\eta}_{\bar{x}}$ and $\gamma(t) : t \mapsto \bar{x} + t\bar{\eta}_{\bar{x}}$ be the tangent line passing through $\bar{x}$. If the solution of $\min_t \bar{f}(\gamma(t))$ is t_*, then the minimum point of $\bar{f}$ restricted on curve ϕ is $\widetilde{\mathcal{R}}_{\bar{x}}(t_*\bar{\eta}_{\bar{x}})$.*

Proof. Let $\bar{x} = (U, R, V)$ and $\bar{\eta}_{\bar{x}} = (\bar{\eta}_U, \bar{\eta}_R, \bar{\eta}_V)$. For any fixed $t > 0$, suppose we have polar decompositions $U + t\bar{\eta}_U = P_1Q_1$ and $V + t\bar{\eta}_V = P_2Q_2$. According to the definition of $\widetilde{\mathcal{R}}$ and the definition of $\bar{f}$ (Equations (23) and (9), respectively),

$$\begin{aligned} \bar{f}(\gamma(t)) &= \bar{f}(\bar{x} + t\bar{\eta}_{\bar{x}}) \\ &= \bar{f}(P_1Q_1, R, P_2Q_2) \\ &= \bar{f}(P_1, Q_1RQ_2^T, P_2) \\ &= \bar{f}(\widetilde{\mathcal{R}}_{\bar{x}}(t\bar{\eta}_{\bar{x}})) = \bar{f}(\phi(t)). \end{aligned} \tag{34}$$

So there is a one-to-one correspondence between values of $\bar{f}$ on curve $\gamma(t)$ and values on $\phi(t)$. Conclusion of the proposition is clear. □

Incorporating this minimizing property into the theoretic framework of the original R3MC, we propose the improved version A-R3MC as shown in Algorithm 2.

Algorithm 2 A-R3MC with New Retraction

Input: Given an initial point $\bar{x}_0 \in \overline{\mathcal{M}}$.

1: Compute Riemannian gradient $\bar{\xi}_i \in \mathcal{H}_{\bar{x}_i} \subset T_{\bar{x}_i}\overline{\mathcal{M}}$.
2: Compute the conjugate direction $\bar{\eta}_i \in \mathcal{H}_{\bar{x}_i}$ by Polak-Ribière+ or Fletcher-Reeves.
3: If $\bar{\eta}_i$ is not a descent direction, reset $\bar{\eta}_i = -\bar{\xi}_i$.
4: Find exact minimum point $\bar{x}_{i+1}$ along $\bar{\eta}_i$ using $\widetilde{\mathcal{R}}$.
5: Repeat until convergence.

In step 2 of Algorithm 2, we can replace the Polak-Ribière+ method by Fletcher-Reeves method in which the factor β in Equation (21) is computed by

$$\beta_i = \frac{\langle \bar{\xi}_i, \bar{\xi}_i \rangle}{\langle \bar{\xi}_{i-1}, \bar{\xi}_{i-1} \rangle}. \tag{35}$$

We investigate the performance of A-R3MC under Fletcher-Reeves and Polak-Ribière+ in Section 5.1.1. In step 3, the reset operation seldom takes place because of the exact line search. In step 4, the exact minimum point $\bar{x}_{i+1}$ is computed according to Proposition 3. Note that we always expect a positive step-size, so the minimization of cost function $\bar{f}(t)$ is done for $t > 0$. Solution t_* of the minimization of this degree 6 polynomial $\bar{f}(t)$ on the positive real line can be efficiently computed numerically. Then retract the vector $t_*\bar{\eta}_{\bar{x}}$ by $\widetilde{\mathcal{R}}_{\bar{x}}$ to find the next iterate $\bar{x}_{i+1} := \widetilde{\mathcal{R}}_{\bar{x}}(t_*\bar{\eta}_{\bar{x}})$.

4.2. Discussion

The minimizing property in Proposition 3 for a retraction is an interesting phenomenon. It is a combination of properties of the cost function, the geometric structure and the optimization method. In designing retractions for a specific problem in the geometric optimization framework, this minimizing property may be taken into consideration.

There is an intuitively similar yet different operation called cascaded update algorithm in a line search as described in [6, Section 4.1]. The cascaded algorithm proceeds in the context of quotient manifold and consists of two updates. The first one minimizes across different equivalence classes while the second "balancing" update is a change of representative along the fiber. The cascaded update algorithm is applicable in quotient manifold structure and aims at providing a better representative among the equivalence classes while our line search adapts an embedding point of view and establishes connections between the tangent line in the ambient space and the searching curve on the embedded manifold through a tailored retraction that provides exact minima along the searching curve.

5. Numerical Experiments

We implemented A-R3MC on basis of the R3MC codes provided by [7]. All experiments were conducted using Matlab on a 2.90 GHz Intel Core i7 machine with 8G RAM.

5.1. Synthetic Data

Testing problems of synthetic data were generated by multiplying three matrices $M = USV^T$ where, U and V are orthogonal bases of two randomly generated matrices according to the standard Gaussian distribution in sizes $n \times r$ and $m \times r$ respectively, and S a diagonal matrix with exponential decaying singular values. In the experiments, CN denotes the condition number of a matrix and OS is the over-sampling factor for a rank-r matrix which is defined as the ratio $|\Omega|/(nr + mr - r^2)$. OS represents the proportion of known entries and these entries are selected from M under uniform probability in the experiments. The initial point $\bar{x}_0 = (U_0, R_0, V_0) \in \overline{\mathcal{M}}$ was generated by multiplying two random matrices $G \in \mathbb{R}^{n\times r}$ and $H^T \in \mathbb{R}^{r\times m}$ firstly, and then performed a SVD to GH^T. We fixed the rank $r = 10$ for the problem matrices and varied the dimension, condition number, and over-sampling ratio in our tests. Unless specified otherwise, all algorithms were terminated if the cost function went below 10^{-10} or the iteration number exceeded the maximum count 500.

5.1.1 Comparing with R3MC

There are two settings for both A-R3MC and R3MC in this comparison. A-R3MC1 and R3MC1 employ the Polak-Ribière+ nonlinear conjugate-gradient method while A-R3MC2 and R3MC2 adopt the Fletcher-Reeves method. In the line search step, all the four algorithms are set to minimize the degree 6 polynomial $\bar{f}(t)$ defined on the tangent line rather than the degree 2 approximation, so computation cost per iteration is about the same.Thus we present the results in terms of iteration numbers as shown in Figure 4.

(a) Ill-conditioning and scaling. Figures 4(a), 4(b) and 4(c) in the first row of Figure 4 are matrix completion problems for small size 5000×5000 while problems in Figures 4(e), 4(f) and 4(g) in the second row are in large-scale $200,000 \times 200,000$. All these problem instances share same OS ratio 3 and same rank 10. From left to right, condition numbers increase from 100 to 10^{10}. We can see that the improvement of A-R3MC1 to R3MC1 is more apparent for large-scale matrices while the improvement of A-R3MC2 to R3MC2 is substantial in all cases.

(b) Low sampling. As observed in [11, Section 5], generally OS > 2 is needed to reliably recover an incoherent matrix after uniform sampling. Figures 4(d) and 4(h) are two cases of low-sampling where the OS ratio is set to a small number 2.1. We see that A-R3MC1 provides a small improvement to R3MC1 and the acceleration of A-R3MC2 over R3MC2 is satisfactory.

(c) Rectangular matrices. Completing a rectangular matrix is more challenging than completing a square matrix. Figure 4(h) takes a rectangular matrix with size $100,000 \times 1000$, OS ratio 2.1 and condition number 100. Both A-R3MC1 and A-R3MC2 show better performance in this example.

(d) Polak-Ribière+ and Fletcher-Reeves nonlinear conjugate-gradient methods. Figure 4 also illustrates interesting behaviors of Polak-Ribière+ (PR) and Fletcher-Reeves (FR) in different scenarios. In these experiments, A-R3MC1 and R3MC1 are equipped with PR while A-R3MC2 and R3MC2 with FR. PR is more efficient when the condition number is small while FR performs better in the ill-conditioned case. A-R3MC2 with the new retraction is the most robust algorithm among the four algorithm settings. The convergence of A-R3MC2 is little affected by the matrix scale or the condition number. In the tested cases, we can see that A-R3MC1 is in general faster than R3MC1 and A-R3MC2 out performs R3MC2.

5.1.2 Comparing with Other Algorithms

We compare A-R3MC1/A-R3MC2 with R3MC, LRGeomCG [11], RTRMC [2], ScGrass-CG [8], and LMaFit [12] as shown in Figure 5. Figure 5(a) are the results of completing a 5000×5000 matrix with condition number 10,000 and Figure 5(b) are completing a $10,000 \times 10,000$ matrix with condition number 10^6. The OS ratios are 3 in these two cases and the maximum iterations are set to be 1000. R3MC is done as in paper [7], *i.e.*, Polak-Ribière+ with accelerated linearized step-size guess. Performances of these algorithms are measured by time in second. The time axis is truncated to a proper length for better visualization. In Figure 5, A-R3MC2 is the only one that converges to the tolerance of 10^{-10} and A-R3MC1 has the second best performance.

5.2. MovieLens Dataset

We compared the algorithms on the MovieLens 10M dataset, which contains the data of 10,000,054 ratings to 10,681 movies by 71,567 users. Ratings of the dataset were made on a 5-star scale with half-star increments and all users rated at least 20 movies. We trained the algorithms by 90 percent of randomly selected data and test them on the remaining 10 percent of data. All tests were repeated 10 times and the values reported were averages. All algorithms took random initializations and stop after 500 iterations, except that RTRMC is stopped when its outer iteration number reaches 200. Both A-R3MC and R3MC employ the Polak-Ribière+ method. In addition, R3MC adapts the accelerated linearized step-size guess. The costs and Mean Square Errors (MSEs) were listed in Table 1. We can see that A-R3MC shows better convergence and competes with other methods in recovery precisions.

6. Conclusion

We proposed a new retraction for the quotient manifold used in R3MC, which is a geometric optimization algorithm for the matrix completion problem. This new retraction possesses a good minimizing property, which provides a way to find the exact minima along the retracted curve on the searching manifold. Incorporating this special retraction into R3MC, we obtained an accelerated version A-R3MC. Extensive numerical experiments showed that A-R3MC outperforms the state-of-the-art geometric algorithms on the more structured synthetic data and is competitive with other algorithms on the real-world data. The interplay between geometry and optimization is interesting and we will investigate deeper into retractions with special properties in our future work.

Acknowledgments

Z. Lin is supported by 973 Program of China (grant no. 2015CB352502), NSF China (grant nos. 61272341 and 61231002), and Microsoft Research Asia Collaborative Research Program. Z. Li would like to thank Bamdev Mishra for providing Matlab codes of R3MC and various benchmarks.

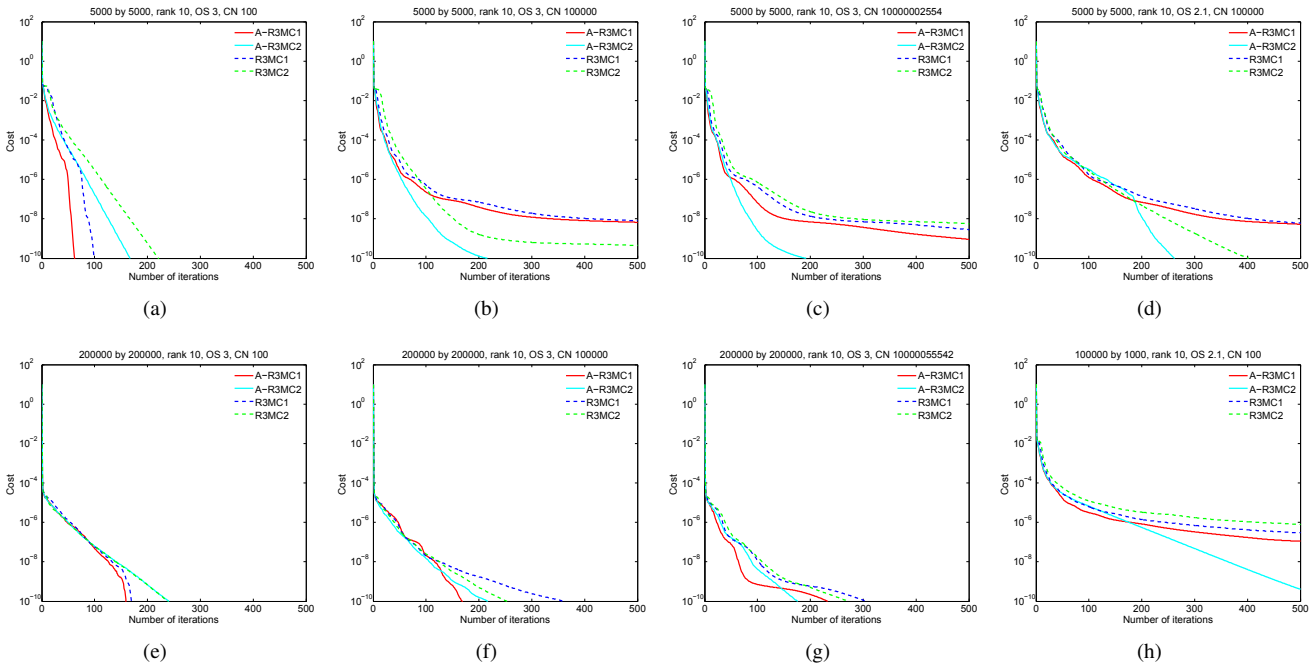

Figure 4. Comparing A-R3MC with R3MC in different scenarios. For the left three figures in the first row, we increased the condition numbers (CNs) while kept other configurations fixed. The corresponding three figures in the second row, in contrast, shows the behaviour of the various algorithms when the matrix sizes became large. Figures 4(d) and 4(h) illustrate the low-sampling case and the rectangular matrix case, respectively.

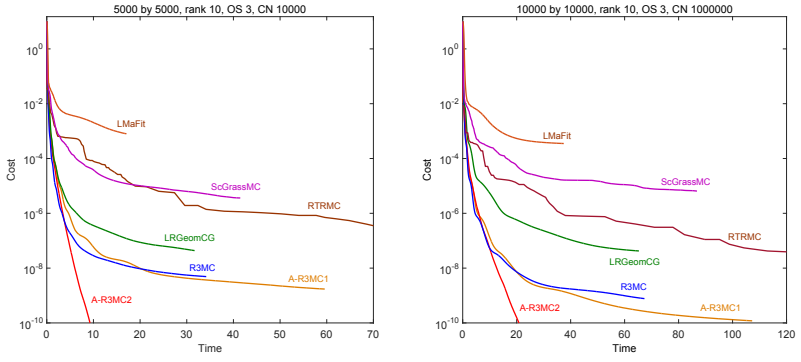

Figure 5. Comparisons of A-R3MC1/A-R3MC2 with R3MC, LRGeomCG, RTRMC, ScGrassMC and LMaFit. Two configurations were shown. For better readability, the lines of RTRMC were pruned.

Rank	A-R3MC	R3MC	LRGeom	RTRMC	ScGrass-CG	LMaFit
6	**0.5725**/0.6573	0.5734/**0.6572**	0.5736/0.6577	0.5854/0.6655	0.5899/0.6626	0.7527/0.7897
7	**0.5573/0.6549**	0.5588/**0.6549**	0.5591/0.6555	0.5776/0.6697	0.5801/0.6639	0.7517/0.7937
8	**0.5441**/0.6543	0.5461/**0.6539**	0.5467/0.6548	0.5703/0.6654	0.5734/0.6633	0.7507/0.7977
9	**0.5330**/0.6548	0.5352/**0.6541**	0.5360/0.6552	0.5651/0.6827	0.5654/0.6652	0.7501/0.8030
10	**0.5228**/0.6561	0.5258/**0.6560**	0.5265/0.6571	0.5793/0.6948	0.5609/0.6696	0.7499/0.8094

Table 1. Costs/MSEs for MovieLens 10M dataset. The MSEs were selected by finding the minimum value in the history of the running process to cope with possible over fitting. A-R3MC has the smallest cost among these methods. The original R3MC performs better in the Test MSEs, but A-R3MC's performance is very close to that of R3MC in most cases.

References

[1] P. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, 2009.

[2] N. Boumal and P.-A. Absil. RTRMC: A Riemannian trust-region method for low-rank matrix completion. In *Advances in Neural Information Processing Systems 24*. 2011.

[3] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

[4] M. P. do Carmo. *Riemannian geometry*. Birkhauser, Boston, 1992.

[5] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235, 2010.

[6] G. Meyer, S. Bonnabel, and R. Sepulchre. Linear regression under fixed-rank constraints: a Riemannian approach. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 2011.

[7] B. Mishra and R. Sepulchre. R3MC: A Riemannian three-factor algorithm for low-rank matrix completion. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 2014.

[8] T. Ngo and Y. Saad. Scaled gradients on Grassmann manifolds for matrix completion. In *Advances in Neural Information Processing Systems 25*. 2012.

[9] J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, pages 713–719. ACM, 2005.

[10] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

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

[12] 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.