CONVERGENCE ANALYSIS OF PROJECTED FIXED-POINT ITERATION ON A LOW-RANK MATRIX MANIFOLD *

D. A. KOLESNIKOV ‡ AND I. V. OSELEDETS‡ ¶

Abstract. In this paper we analyse convergence of projected fixed-point iteration on a Riemannian manifold of matrices with fixed rank. As a retraction method we use "projector splitting scheme". We prove that the projector splitting scheme converges at least with the same rate as standard fixed-point iteration without rank constraints. We also provide counter-example to the case when conditions of the theorem do not hold. Finally we support our theoretical results with numerical experiments.

Key words. fixed-point iteration, Riemannian optimization framework, low-rank approximation

AMS subject classifications. 93B40, 58C30, 47J25, 65F30.

1. Introduction. In many applications it is well-known that the solution of the optimization problem can be approximated by low-rank matrices or tensors, i.e. it lies on a certain manifold [2, 1]. Thus, instead of minimizing the full functional, the framework of Riemannian optimization can be very effective in terms of storage [13, 10]. There are different approaches for the optimization over low-rank manifolds, including projection onto the tangent space [9] conjugate-gradient type methods [12], second-order methods [4]. The manifolds of matrices with bounded ranks and tensors with fixed tensor train and hierarchical ranks are of crucial importance in many high-dimensional problems, and are examples of Riemannian manifolds with a very particular polylinear structure. In this paper we consider the two-dimensional (matrix) case and study the convergence of the projected gradient-type methods and show that if the original method converges, its manifold version based on the so-called *projector-splitting method* is guaranteed to converge at least with the same rate and some additional conditions on the initial approximation. This is up to a certain extent an unexpected result, since the standard estimates include the curvature of the manifold. For the manifold of matrices of rank r, the curvature is given by $1/\sigma_{\min}$, i.e. if the matrix is close to the matrix of a smaller rank, such estimates are useless in practice. Our results show that the curvature is not important for the convergence.

Consider an iterative process

$$X_{k+1} = \Phi(X_k), \quad k = 0, \dots \tag{1.1}$$

where $Y_k \in \mathbb{R}^{n\times m}$ and Φ is a contraction with parameter δ. Then, X_k converges linearly to X_*, for $k \to \infty$, i.e.

$$\|X_{k+1} - X_*\| \leq \delta \|X_k - X_*\|,$$

for some matrix norm $\|\cdot\|$. Also we assume that the initial point and the final points are on the manifold, i.e. and

$$X_0, X_* \in \mathcal{M}_r, \quad \mathcal{M}_r = \left\{ X \middle| \operatorname{rank} X \leq r \right\}.$$

*This work was supported by Russian Science Foundation grant 14-1100659

‡Skolkovo Institute of Science and Technology, Novaya St. 100, Skolkovo, Odintsovsky district, 143025 Moscow Region, Russia (denis.kolesnikov@skoltech.ru, i.oseledets@skoltech.ru)

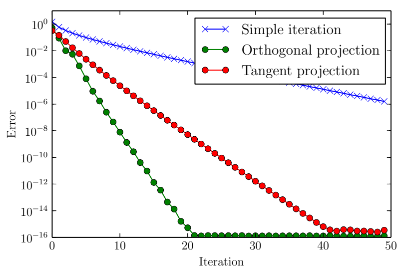
¶Institute of Numerical Mathematics, Gubkina St. 8, 119333 Moscow, Russia

From (1.1) we create the projected version as

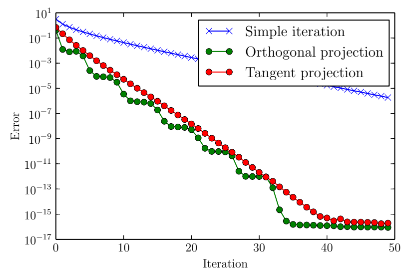
$$Y_{k+1} = I(Y_k, \Phi(Y_k) - Y_k), \quad k = 0, \ldots, \tag{1.2}$$

where $I(Z, H)$ is the *projector-splitting integrator* [9] which is known to be a *retraction* to the manifold [4]. There are many other possible choices for the retraction, but in this paper we consider only one of them and all the convergence estimates are proven for the method (1.2).

Our approach is based on the splitting the error $\|X_k - X_*\|$ into two components. The first component is a projection on the tangent space of the manifold at some intermediate point and shows how close current point to stationary point in the sense of Riemannian metric on the manifold. The second component is the projection on normal space at the same point and is related to the manifold curvature. The typical case convergence is presented at Figure 1.1a. However, much more interesting pattern is possible. See Figure 1.1b.



(a) Typical case convergence.



(b) Stair case convergence.

In both cases, although the curvature influences only on but the convergence is not worse than for the full case.

2. Projector-splitting integrator. The projector-splitting integrator was originally proposed [6] as an integration scheme for the equations of motions of dynamical low-rank approximation. However, the only information it requires, are two matrices, A_0, A_1, at subsequent time steps. Thus it is very natural to consider it for the discrete time problems, and moreover, it can be formally viewed as a retraction onto the manifold of rank-r matrices. It is formulated as follows.

Given a rank-r matrix in the form $A_0 = U_0 S_0 V_0^\top, U_0^\top U_0 = V_0^\top V_0 = I_r$ and a direction D, it provides the retraction of $A_0 + D$ back onto the manifold by the following steps:

Algorithm 1: The projector splitting retraction

Data: $A_0 = U_0 S_0 V_0^\top, \quad D$
Result: $A_1 = U_1 S_1 V_1^\top$
$U_1, S' = \mathrm{QR}(U_0 S_0 + D V_0)$;
$S'' = S' - U_1^\top D V_0^\top$;
$V_1, S_1^\top = \mathrm{QR}(V_0 S''^\top + D^\top U_1)$;

Note that the QR-factorizations in the intermediate steps are non-unique, but the final result $U_1 S_1 V_1^\top$ does not depend on it. For the details we refer the reader to [9].

We will denote the result of Algorithm 1 as $I(A_0, D)$. Define $\mathcal{T}(X)$ as the tangent space of $X \in \mathcal{M}_r$ The following Lemma provides a new interpretation of the projector-splitting integrator as a projection onto the tangent plane in some intermediate point.

Lemma 2.1. *Let* $\operatorname{rank} A_0 = U_0 S_0 V_0^\top, \quad U_0^\top U_0 = V_0^\top V_0 = I_r$, $D \in \mathbb{R}^{n\times m}$. *Then,*

$$I(A_0, D) = P_{\mathcal{T}(X)}(A_0 + D), \quad I(A_0, D), A_0 \in \mathcal{T}(X). \tag{2.1}$$

where X is some matrix of rank r.

Proof. It is sufficient to select $X = U_1 S V_0^\top$ for any non-singular S, and U_1 is defined as in the Algorithm (1). Note from the construction, that both the initial and the final points lie in the tangent space $\mathcal{T}(X)$. ▯

3. Decomposition of the error into the normal and tangent parts. Let us write one step of the iterative process (1.2) as

$$Y_1 = I(Y_0, \Phi(Y_0) - Y_0). \tag{3.1}$$

Using the projector form (2.1) we have

$$Y_1 = P_{\mathcal{T}(X)}(\Phi(Y_0)),$$

and the error can be written as

$$E_1 = Y_1 - X_* = P_{\mathcal{T}(X)}(\Phi(Y_0) - \Phi(X_*)) + P_{\mathcal{T}(X)}(X_*) - X_*. \tag{3.2}$$

Due to the contraction property we can bound

$$\|\Phi(Y_0) - \Phi(X_*)\| \leq \delta \|E_0\|.$$

It is natural to introduce the notation

$$P_{\mathcal{T}(X)}(X_*) - X_* = -P^{\perp}_{\mathcal{T}(X)}(X_*),$$

since it is the normal to the tangent space component of X_* at point X. Thus the error at the next step satisfies

$$\varepsilon_1^2 = \|E_1\|^2 = \varepsilon_\tau^2 + \varepsilon_\perp^2.$$

From the definition it is easy to see that

$$\varepsilon_\tau = \|P_{\mathcal{T}(X)}(\Phi(Y_0) - \Phi(X_*))\| \leq \|\Phi(Y_0) - \Phi(X_*)\| \leq \delta \varepsilon_0.$$

The estimate for the decay of $\varepsilon_\perp = \|P^{\perp}_{\mathcal{T}(X)}(X_*)\|$ is much less trivial.

4. Estimate for the normal component of the error. From the definition of the error we have

$$\Phi(Y_0) = X_* + H,$$

and $\|H\| \leq \delta \varepsilon_0$. Since Y and X_* are on the manifold, they admit factorizations

$$Y = U_0 S_0 V_0^\top, \quad X_* = U_* S_* V_*^\top,$$

where U_*, V_*, U_0 and V_0 are orthonormal. If ε_0 is small, one can expect that the subspaces spanned by columns of V_0 and V_* are close; however, the estimates depend on the smallest singular values of X_*. The following Theorem gives a bound on the normal component.

THEOREM 4.1. *Let $X_* = U_* S_* V_*^\top$, where $V_*^\top V_* = U_*^\top U_* = I_q, \quad q \le r$ and H is an $n \times m$ matrix, V_0 be an $m \times r$ matrix with orthonormal columns and U_1 be any orthogonal basis for the column space of the matrix $(X_* + H)V_0$. Then, the norm of $P^\perp(X_*)$ defined as*

$$P^\perp(X_*) = (I - U_1 U_1^\top) X_* (I - V_0 V_0^\top). \tag{4.1}$$

can be bounded as

$$\|P^\perp(X_*)\| \le \|H\| \| \tan \angle(V_0, V_*)\|. \tag{4.2}$$

Proof. First, we find an $r \times r$ orthonormal matrix Q such that

$$\Psi Q = (V_*^\top V_0) Q = \begin{bmatrix} \widehat{\Psi} & 0_{r-q} \end{bmatrix}, \tag{4.3}$$

where matrix $\widehat{\Psi}$ has size $q \times q$. Since the multiplication by the orthogonal matrix Q does not change the projector

$$V_0 V_0^\top = (V_0 Q)(V_0 Q)^\top,$$

we can always assume that the matrix Ψ is already in the form (4.3). Since U_1 spans the columns space of $(X_* + H)V_0$, we have

$$(U_1 U_1^\top)(X_* V_0 + H V_0) = X_* V_0 + H V_0. \tag{4.4}$$

From this equation we have

$$X_* V_0 = U_1 U_1^\top X_* V_0 + U_1 U_1^\top H V_0 - H V_0 = U_* S_* V_*^\top V_0 = U_* \begin{bmatrix} \widehat{\Psi} & 0 \end{bmatrix}. \tag{4.5}$$

Introduce the matrix $V_0^{(q)}$ comprised of the first q column of the matrix V_0. From (4.5) we have

$$U_* S_* \widehat{\Psi} = U_1 U_1^\top X_* V_0^{(q)} + U_1 U^\top H V_0^{(q)} - H V_0^{(q)}.$$

Thus,

$$U_* S_* = U_1 \Psi_1 - H V_0^{(q)} \widehat{\Psi}^{-1}, \tag{4.6}$$

Note, that

$$\|P^\perp\| = \|(I - U_1 U_1^\top) X_* (I - V_0 V_0^\top)\| \le \|(I - U_1 U_1^\top) X_* (I - V_0^{(q)} V_0^{(q)})^\top))\|,$$

and from (4.4) it follows also that

$$(I - U_1 U_1^\top)(X_* + H) V_0^{(q)} (V_0^{(q)})^\top = 0.$$

For simplicity, denote

$$P_q^\perp(X_*) = (I - U_1 U_1^\top) X_* (I - V_0^{(q)} (V_0^{(q)})^\top).$$

Then,

$$P_q^{\perp}(X_*) = (I - U_1U_1^\top)X_* - (I - U_1U_1^\top)X_*V_0^{(q)}(V_0^{(q)})^\top = \\ = (I - U_1U_1^\top)X_* + (I - U_1U_1^\top)HV_0^{(q)}(V_0^{(q)})^\top. \tag{4.7}$$

Replacing U_*S_* in (4.7) by (4.6) we get

$$\begin{aligned} P_{\mathcal{T}(X)}^{\perp} &= (I - U_1U_1^\top)U_*V_*^\top + (I - U_1U_1^\top)HV_0^{(q)}(V_0^{(q)})^\top) \\ &= (I - U_1U_1^\top)HV_0^{(q)}(V_0^{(q)})^\top - (I - U_1U_1^\top)HV_0^{(q)}\widehat{\Psi}^{-1}V_*^\top \\ &= (I - U_1U_1^\top)HV_0^{(q)}(V_0^{(q)})^\top - \widehat{\Psi}^{-1}V_*^\top). \end{aligned} \tag{4.8}$$

To estimate the norm, note that

$$\|P_{\mathcal{T}(X)}^{\perp}\| \leq \|H\|\|(V_0^{(q)})^\top - \widehat{\Psi}^{-1}V_*^\top\|.$$

Introduce the matrix

$$B = (V_0^{(q)})^\top - \widehat{\Psi}^{-1}V_*^\top.$$

We have

$$\|X_*(I - V_0V_0^\top)\| = \|(X_* - Y_0)(I - V_0V_0^\top)\| \leq \|X_* - Y_0\|.$$

Replacing X_* by $U_*S_*V_*^\top$ we have

$$\|U_*(V_*^\top - \Psi V_0^\top)\| = \|U_*(V_*^\top - \widehat{\Psi}(V_0^{(q)})^\top)\|.$$

Thus,

$$\|V_*^\top - \widehat{\Psi}(V_0^{(q)})^\top\| \leq \frac{\|X_* - Y_0\|}{\sigma_q}.$$

Introduce the matrix $C = V_*^\top - \widehat{\Psi}(V_0^{(q)})^\top$. Then,

$$\|C\|^2 = \|CC^\top\| = \|I - \widehat{\Psi}\widehat{\Psi}^\top\| \leq \frac{\|X_* - Y_0\|^2}{\sigma_q^2}.$$

Then, we have

$$\sin\theta \leq \frac{\|X_* - Y_0\|}{\sigma_q},$$

whereas we require to bound

$$\tan\theta = \frac{\sin\theta}{\sqrt{1 - \sin^2\theta}}.$$

Let $\widehat{\Psi} = U\Lambda V^\top$ be the singular value decomposition of $\widehat{\Psi}$. From the definition of the angles between subspaces we have

$$\Lambda = \cos\angle(V_*^\top, V_0^{(q)}) = \cos\angle(V_*^\top, V_0),$$

therefore

$$\|B\|^2 = \|\cos^{-2}\angle(V_*^\top, V_0) - 1\| = \|\tan^2\angle(V_*^\top, V_0)\|,$$

which completes the proof. □

5. Error estimate. Theorem 4.1 shows that the normal component can decay as a tangent component squared. Unfortunately, convergence of the projector splitting method in general is not guaranteed. In section 6 we give the example for which sequence Y_k converges to a matrix different from X_*. In this section we derive sufficient conditions for convergence of projector splitting method.

We consider one step of the projector splitting scheme.

LEMMA 5.1. *Let us denote the initial point $Y_0 = U_0 S_0 V_0^\top$, the next step point $Y_1 = U_1 S_1 V_1^\top$ and the fixed point $X_* = U_* S_* V_*^\top$. We assume that S_* is a diagonal matrix:*

$$S_* = \sum_{k=1}^{r} s_k e_k e_k^\top,$$

where s_k is the k-singular value and e_k is the corresponding vector from the standard basis. Let us denote

$$\cos^2 \phi_{Li,k} = \|U_i U_i^\top U_* e_k\|_F^2, \qquad \cos^2 \phi_{Ri,k} = \|e_k^\top V_*^\top V_i V_i^\top\|_F^2,$$
$$\sin^2 \phi_{Li,k} = \|(I - U_i U_i^\top) U_* e_k\|_F^2, \qquad \sin^2 \phi_{Ri,k} = \|e_k^\top V_*^\top (I - V_i V_i^\top)\|_F^2.$$

Assume that

$$\delta^2 \|Y_0 - X_*\|_F^2 + \sum_{k=1}^{r} s_k^2 \sin^2 \phi_{R0,k} \le s_r. \tag{5.1}$$

Then the next inequality holds:

$$\|Y_1 - X_*\|_F^2 \le \delta^2 \|Y_0 - X_*\|_F^2 +$$
$$+\Big(\delta^2 \|Y_0 - X_*\|_F^2 - \sum_{k=1}^{r} s_k^2 \sin^2 \phi_{R1,k}\Big) \frac{\sum\limits_{k=1}^{r} s_k^2 \sin^2 \phi_{R0,k}}{s_r - \sum\limits_{k=1}^{r} s_k^2 \sin^2 \phi_{R0,k} - \sum\limits_{k=1}^{r} s_k^2 \sin^2 \phi_{R1,k}}. \tag{5.2}$$

Proof. Without the loss of generality we can assume that

$$U_1 = \begin{bmatrix} I_r \\ 0_{(n-r)\times r} \end{bmatrix}, \quad V_0 = \begin{bmatrix} I_r \\ 0_{(m-r)\times r} \end{bmatrix}.$$

Then we use the following block representation of $Y_0, \Phi(Y_0), Y_1$ and X_*:

$$Y_0 = U_0 S_0 V_0^\top = \begin{bmatrix} D_1^0 & 0 \\ D_3^0 & 0 \end{bmatrix}, \Phi(Y_0) = \begin{bmatrix} D_1^1 & D_2^2 \\ 0 & D_4^1 \end{bmatrix},$$
$$Y_1 = U_1 S_1 V_1^\top = \begin{bmatrix} D_1^1 & D_2^1 \\ 0 & 0 \end{bmatrix}, X_* = U_* S_* V_*^\top = \begin{bmatrix} E_1 & E_2 \\ E_3 & E_4 \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\|Y_1 - X_*\|_F^2 &= \|D_1^1 - E_1\|_F^2 + \|D_2^1 - E_2\|_F^2 + \|E_3\|_F^2 + \|E_4\|_F^2 \le \\ &\le \big(\|D_1^1 - E_1\|_F^2 + \|D_2^1 - E_2\|_F^2 + \|E_3\|_F^2 + \|D_4^1 - E_4\|_F^2\big) + \big(\|E_4\|_F^2\big) = \\ &= \|\Phi(Y_0) - X_*\|_F^2 + \|(I - U_1 U_1^\top) X_* (I - V_0 V_0^\top)\|_F^2 \le \\ &\le \delta^2 \|Y_0 - X_*\|_F^2 + \|(I - U_1 U_1^\top) X_* (I - V_0 V_0^\top)\|_F^2.\end{aligned}$$

We want to estimate $\|(I - U_1U_1)X_*(I - V_0V_0)\|_F^2$. For that purpose we exploit contraction property of Φ:

$$
\begin{aligned}
\|U_1U_1^\top(\Phi(Y_0) - X_*)\|_F^2 + \|(I - U_1U_1^\top)(\Phi(Y_0) - X_*)\|_F^2 = \\
= \|(\Phi(Y_0) - X_*)\|_F^2 \le \delta^2\|Y_0 - X_*\|_F^2, \\
\|U_1U_1^\top(X_*)(I - V_1V_1^\top)\|_F^2 + \|(I - U_1U_1^\top)(X_*)V_0V_0^\top\|_F^2 \le \delta^2\|Y_0 - X_*\|_F^2, \\
\|(I - U_1U_1^\top)(X_*)V_0V_0^\top\|_F^2 - \|(I - U_1U_1^\top)(X_*)(I - V_1V_1^\top)\|_F^2 \le \\
\le \delta^2\|Y_0 - X_*\|_F^2 - \|(X_*)(I - V_1V_1^\top)\|_F^2.
\end{aligned}
$$

Then the inequality (5) transforms to

$$
\begin{aligned}
&\sum_{k=1}^{r} s_k^2\|(I - U_1U_1^\top)U_*e_1\|_F^2\|e_k^\top V_0V_0^\top\|_F^2 - \\
&- \sum_{k=1}^{r} s_k^2\|(I - U_1U_1^\top)U_*e_k\|\|e_k^\top V_*^\top(I - V_1V_1^\top)\|_F^2 \le \\
&\le \delta^2\|Y_0 - X_*\|_F^2 - \sum_{k=1}^{r} s_k^2\|U_*e_k\|_F^2\|e_k^\top V_*^\top(I - V_1V_1^\top)\|_F^2.
\end{aligned}
\tag{5.3}
$$

Using (**??**) we have

$$
\sum_{k=1}^{r} \sin^2\phi_{L1,k} s_k^2(\cos^2\phi_{R0,k} - \sin^2\phi_{R1,k}) \le \delta^2\|Y_0 - X_*\|_F^2 - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R1,k}.
$$

Inequality (5.1) guarantees that

$$
\begin{gathered}
\sum_{k=1}^{r} s_k^2\sin^2\phi_{R0,k} - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R1,k} < s_r^2, \\
0 < \max_{1\le k\le r}\left(\cos^2\phi_{R0,k} - \sin^2\phi_{R1,k}\right).
\end{gathered}
$$

Therefore

$$
\begin{aligned}
&\sum_{k=1}^{r} s_k^2\sin^2\phi_{L1,k}\sin^2\phi_{R0,k} \le \\
&\le \left(\delta^2\|Y_0 - X_*\|_F^2 - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R1,k}\right)\max_{1\le k\le r}\frac{\sin^2\phi_{R0,k}}{\cos^2\phi_{R0,k} - \sin^2\phi_{R1,k}} \le \\
&\le \left(\delta^2\|Y_0 - X_*\|_F^2 - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R1,k}\right)\frac{\sum_{k=1}^{r} s_k^2\sin^2\phi_{R0,k}}{s_r^2 - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R0,k} - \sum_{k=1}^{r} s_k^2\sin^2\phi_{R1,k}}.
\end{aligned}
$$

i.e. (5.2) is proven. □

For convenience we introduce new variables:

$$s = \delta^2, \quad p_k = \frac{\|Y_k - X_*\|_F^2}{s_r^2}, \quad q_k = \frac{1}{s_r^2} \sum_{k=1}^{r} s_k^2 \sin^2 \phi_{Rk}, \tag{5.4}$$

Now we can formulate the connection between the subsequent steps:

$$p_{k+1} \le sp_k + \frac{(sp_k - q_{k+1})q_k}{1 - q_k - q_{k+1}}, \quad 0 \le q_{k+1} \le sp_k. \tag{5.5}$$

We can derive upper estimate for p_k:

THEOREM 5.2. *Assume that* $0 < s < 1$, $0 \le q_0 \le 1$, $0 < p_0$.
Consider $p_k, q_k, k \in \mathbf{N}$ *that satisfy* (5.4). *Assume that* $4\frac{p_0}{(1-q_0)^2}\frac{s}{1-s} < 1$. *Then the next inequalities hold:*

$$p_k \le \frac{p_0}{c_*(s, p_0, q_0)} s^k, \quad 0 < c_*(s, p_0, q_0) \le 1 - \sum_{j=0}^{k} q_k \le sp_{k-1} + q_{k-1}, \tag{5.6}$$

where

$$c_*(s, p_0, q_0) = \frac{p_0}{1 - q_0} \frac{s}{1 - s} \left(\frac{2}{1 + \sqrt{1 - 4\frac{p_0}{(1-q_0)^2}\frac{s}{1-s}}} \right).$$

Proof. The parameter $c_*(s, p_0, q_0)$ is the positive solution of the equation:

$$c_*(s, p_0, q_0) = 1 - q_0 - p_0 \frac{s}{1 - s} \frac{1}{c_*(s, p_0, q_0)}.$$

We will use mathematical induction to prove (5.6). The base case follows from $0 < c_*(s, p_0, q_0) < 1$

$$p_0 \le \frac{p_0}{c_*(s, p_0, q_0)}, \quad c_*(s, p_0, q_0) \le 1 - q_0.$$

Consider the inductive step. Assume that (5.6) holds for every $i < k$ for some k. Then,

$$\begin{aligned} p_{k+1} &\le sp_k + \frac{(sp_k - q_{k+1})q_k}{1 - q_k - q_{k+1}} = \\ &= sp_k \frac{1 - q_k}{1 - q_k - q_{k+1}} - \frac{q_{k+1}q_k}{1 - q_k - q_{k+1}} \le s \frac{p_k}{1 - \frac{q_{k+1}}{1-q_k}}. \end{aligned} \tag{5.7}$$

We can expect that the term $\frac{q_{k+1}q_k}{1 - q_k - q_{k+1}}$ is sufficiently smaller than the p_{k+1} and decays as p_{k+1}^2 due to $q_k \sim p_k$. Finally,

$$p_{k+1} \le \frac{sp_k}{1 - \left(\frac{q_{k+1}}{1 - q_k}\right)} \le \frac{s^{k+1} p_0}{\prod_{j=0}^{k} \left(1 - \frac{q_{j+1}}{1 - q_j}\right)}. \tag{5.8}$$

It is easy to prove that in the case $\sum\limits_{j=0}^{k} q_j < 1$ we have

$$\prod_{j=0}^{k}\left(1-\frac{q_{j+1}}{1-q_j}\right) \le 1-\sum_{j=0}^{k+1} q_k.$$

It leads to

$$p_{k+1} \le \frac{s^{k+1}p_0}{1-\sum\limits_{j=0}^{k} q_j} \le \frac{s^{k+1}p_0}{c_*(s,p_0,q_0)},$$

therefore

$$c_*(s,p_0,q_0) = 1-q_0-\frac{p_0}{c_*(s,p_0,q_0)}\frac{s}{1-s} = 1-q_0-s\sum_{k=0}^{\infty}\frac{p_0}{c_*(s,p_0,q_0)}s^i \le$$
$$\le 1-q_0-s\sum_{j=0}^{k+1}p_j \le 1-\sum_{j=0}^{k+1}q_j \le 1-q_k-sp_k.$$

The inductive step is proven. □

The final estimate is

$$p_n \le \frac{p_0}{c_*(s,p_0,q_0)}s^n = \frac{p_0}{1-q_0}s^n\left(\frac{1+\sqrt{1-4\dfrac{p_0}{(1-q_0)^2}\dfrac{s}{1-s}}}{2\dfrac{p_0}{(1-q_0)^2}\dfrac{s}{1-s}}\right).$$

Note that if the condition $4\dfrac{p_0}{(1-q_0)^2}\dfrac{s}{1-s} < 1$ does hold, then the condition $sp_0+q_0<1$ does hold as well.

Corollary 5.3. *Define Y_k as in* (1.2), *X_*, s_k and $\sin^2\phi_{R0,k}$ as in Lemma 5.1. Assume that the next inequality holds*

$$4\frac{\|Y_0-X_*\|}{\left(s_r^2-\sum\limits_{k=1}^{r}s_k^2\sin^2\phi_{R0,k}\right)^2} < 1.$$

Then the sequence Y_k converges to X_* and the following inequality holds

$$\|Y_k-X_*\| < c(\delta,Y_0,X_*)\|Y_0-X_*\|\delta^k,$$

where

$$c(\delta,Y_0,X_*) = \frac{1+\sqrt{1-4\dfrac{\delta^2\|Y_0-X_*\|^2}{(1-\delta^2)\left(s_r^2-\sum\limits_{k=1}^{r}s_k^2\sin^2\phi_{R0,k}\right)^2}}}{2\dfrac{\delta^2\|Y_0-X_*\|^2}{(1-\delta^2)\left(s_r^2-\sum\limits_{k=1}^{r}s_k^2\sin^2\phi_{R0,k}\right)^2}}.$$

This estimate guarantees if the initial point is close enough to the fixed point then the projector splitting method in the worst case has the same convergence rate as the fixed-point iteration method. Also the estimate requires that the distance between the initial point and the fixed point is less than the smallest singular value of the fixed point s_r. In the next section we give the example for which this condition do not hold and the projector splitting method does not converges to the true solution.

6. Counter-example. Consider the case $n = 2$, $r = 1$. We will need the following auxiliary result:

LEMMA 6.1. *Let the mapping* $\Phi : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ *be defined as*

$$\begin{aligned} \Phi(Y) &= X_* + \delta \|Y - X_*\|_F X_\perp, \\ X_* &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, X_\perp = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{6.1}$$

Let us consider δ, d_*, q_{max} *and* s *that satisfy*

$$\begin{aligned} &0 < \delta < 1, \quad 1 < \delta^2 + \delta^6, \quad d_* = \frac{1}{\sqrt{1-\delta^2}}, \\ &0 < q_{max}, \quad 0 < s, \quad \frac{1}{\delta^4 d_*^2}\Big(1 + \frac{q_{max}}{\delta^2 d_*^2}\Big) \le \delta^2 - \frac{s}{\delta^2 d_*^2}. \end{aligned} \tag{6.2}$$

Denote the set

$$\Omega = \left\{ \{p, q\} \middle| 0 \le p, \quad 0 \le q \le q_{max}, \quad \frac{q}{p} \le s \right\}$$

and the function

$$f : \Omega \to \mathbb{R}^2_{0,+}, \quad f(\{p, q\}) = \left\{ \frac{1 + \delta^2 p}{1 + \dfrac{q}{\delta^2 d_*^2 (1+p)}} - 1, \quad \frac{q}{\delta^4 d_*^2 (1+p)} \right\},$$

and

$$f^{*n}(x) = \underbrace{f(\dots f(x)\dots)}_{n \text{ times}}.$$

Then

$$f(\Omega) \subset \Omega, \quad \forall x \in \Omega, \quad \lim_{n\to\infty} f^{*n}(x) = \{0, 0\}.$$

Proof. It is important to note that $1 < \delta^4 d_*^2 = \dfrac{\delta^4}{1-\delta^2}$ because of the choice of δ(6.1). Let us denote $f(\{p, q\}) = \{p_1, q_1\}$. Then

$$q_1 = \frac{q}{\delta^4 d_*^2 (1+p)} \le \frac{q}{\delta^4 d_*^2} < q \le q_{max}, \tag{6.3}$$

and therefore

$$
\begin{aligned}
\frac{q_1}{q} &= \frac{1}{\delta^4 d_*^2(1+p)} \le \frac{1}{\delta^4 d_*^2},\\
\frac{p_1}{p} &= \frac{1}{p}\Big(d\frac{1+\delta^2 p}{1+\dfrac{q}{\delta^2 d_*^2(1+p)}} - 1\Big) \ge \frac{1}{p}\Big(\frac{1+\delta^2 p}{1+\dfrac{q}{\delta^2 d_*^2}} - 1\Big) =\\
&= \Big(\delta^2 - \frac{1}{\delta^2 d_*^2}\frac{q}{p}\Big)/\Big(1+\frac{q}{\delta^2 d_*^2}\Big) \ge \Big(\delta^2 - \frac{s}{\delta^2 d_*^2}\Big)/\Big(1+\frac{q_{max}}{\delta^2 d_*^2}\Big) \ge \frac{1}{\delta^4 d_*^2}.
\end{aligned}
$$

Finally we have

$$
\frac{q_1}{p_1} \le \frac{q/\delta^4 d_*^2}{p/\delta^4 d_*^2} = \frac{q}{p} \le s. \tag{6.4}
$$

The statement $f(\Omega) \subset \Omega$ follows from (6.3) and (6.4). Also the following inequalities hold

$$
\frac{p_1}{p} = \frac{1}{p}\left(\frac{1+\delta^2 p}{1+\dfrac{q}{\delta^2 d_*^2(1+p)}} - 1\right) \le \delta^2, \quad \frac{q_1}{q} \le \frac{1}{\delta^4 d_*^2}.
$$

The inequalities (6) guarantee linear convergence of $f^{*n}(x)$ to $\{0,0\}$ for every $x \in \Omega$. ▯

LEMMA 6.2. *Let contraction mapping Φ is defined as in lemma 6.1. Let us consider parameters δ, d_*, contraction mapping Φ and the set Ω and the function f that satisfy condition of lemma 6.1. Let us denote the set of rank-1 2×2 real matrices $M_{2,1}(\mathbf{R})$, $\phi_R(X)$ - right angle for rank-1 2×2 matrix X and*

$$
\begin{aligned}
&\mathcal{M}_{2,1}^{'} = \Big(X | X \in M_{2,1}(\mathbf{R}), \sin^2(\phi_R(X)) > 0\Big),\\
&\pi : \mathcal{M}_{2,1}^{'} \to \mathbf{R}_{0,+}^2, \quad \pi(X) = \left\{\frac{\|X - X_*\|_F^2}{d_*^2} - 1, \quad \operatorname{ctg}^2\phi_R(X)\right\}.
\end{aligned} \tag{6.5}
$$

Assume that

$$
\begin{aligned}
Y_0 &= (\cos\phi_{L0} \quad \sin\phi_{L0})\, s_0 \begin{pmatrix}\cos\phi_{R0}\\ \sin\phi_{R0}\end{pmatrix} \in \pi^{-1}(\Omega),\\
Y_1 &= I(Y_0, \Phi(Y_0) - Y_0) = (\cos\phi_{L1} \quad \sin\phi_{L1})\, s_0 \begin{pmatrix}\cos\phi_{R1}\\ \sin\phi_{R1}\end{pmatrix}.
\end{aligned} \tag{6.6}
$$

Then the following equalities hold

$$
\pi(Y_1) = f(\pi(Y_0)), \quad \operatorname{ctg}^2\phi_{L1} < \operatorname{ctg}^2\phi_{R1}. \tag{6.7}
$$

Proof. We will use the equivalent form of Algorithm 1

$$
\begin{aligned}
U_1, S' &= \mathrm{QR}((A_0 + D)V_0),\\
V_1, S_1^\top &= \mathrm{QR}((A_0 + D^\top)U_1).
\end{aligned} \tag{6.8}
$$

Let us consider $Y_0 = U_0 S_0 V_0^\top$, $d_0 = \|Y_0 - X_*\|_F$ and $V_0 = \begin{pmatrix}\cos\phi_{R0}\\ \sin\phi_{R0}\end{pmatrix}$. Then

$$\begin{aligned}
&\Phi(Y_0) = \begin{pmatrix}1 & 0\\ 0 & \delta d_0\end{pmatrix}, \quad U_1, S' = \mathrm{QR}\left(\begin{pmatrix}\cos\phi_{R0}\\ \delta d_0 \sin\phi_{R0}\end{pmatrix}\right),\\
&V_1, S_1^\top = \mathrm{QR}\left(\frac{1}{\sqrt{1 + (\delta^2 d_0^2 - 1)\sin^2\phi_{R0}}}\begin{pmatrix}\cos\phi_{R0}\\ \delta^2 d_0^2 \sin\phi_{R0}\end{pmatrix}\right).
\end{aligned} \tag{6.9}$$

Finally we get:

$$U_1 S_1 V_1^\top = \begin{pmatrix}\cos\phi_{R0}\\ \delta d_0 \sin\phi_{R0}\end{pmatrix} \frac{1}{1 + (\delta^2 d_0^2 - 1)\sin^2\phi_{R0}} \begin{pmatrix}\cos\phi_{R0} & \delta^2 d_0^2 \sin\phi_{R0}\end{pmatrix} \tag{6.10}$$

It is important to note that $\cos^2\phi_{L1} < \cos^2\phi_{R1} < \cos^2\phi_{R0}$ in case $1 < \delta d_*$ (and our choice of δ provides that). The equality (6.10) guarantees if $0 < \sin^2\phi_{R0}$ then $0 < \sin^2\phi_{R1}$. So

$$\begin{aligned}
d_1^2 &= S_1^2 + \left(1 - \frac{\cos^2\phi_{R0}}{\cos^2\phi_{R0} + \delta^2 d_0^2 \sin^2\phi_{R0}}\right)^2 - \left(\frac{\cos^2\phi_{R0}}{\cos^2\phi_{R0} + \delta^2 d_0^2 \sin^2\phi_{R0}}\right)^2 =\\
&= \frac{\cos^2\phi_{R0} + \delta^4 d_0^4 \sin^2\phi_{R0}}{\cos^2\phi_{R0} + \delta^2 d_0^2 \sin^2\phi_{R0}} + 1 - \frac{2\cos^2\phi_{R0}}{\cos^2\phi_{R0} + \delta^2 d_0^2 \sin^2\phi_{R0}} = \frac{1 + \delta^2 d_0^2}{1 + \mathrm{ctg}^2\phi_{R0}/(\delta^2 d_0^2)}
\end{aligned} \tag{6.11}$$

Let us denote $p_0 = d_0^2/d_*^2 - 1$ and $q_0 = \mathrm{ctg}^2(\phi_{R0})$. Then

$$\begin{aligned}
&\frac{d_1^2}{d_*^2} - 1 = \frac{1}{d_*^2}\left(\frac{1 + \delta^2 d_0^2}{1 + \mathrm{ctg}^2\phi_R/(\delta^2 d_0^2)}\right) - 1 =\\
&\frac{1}{d_*^2}\left(\frac{1 + \delta^2 d_0^2}{1 + q_0/(\delta^2 d_0^2)}\right) - 1 = \frac{1 + \delta^2 p_0}{1 + \frac{q_0}{\delta^2 d_*^2 (1 + p_0)}} - 1,\\
&q_1 = \mathrm{ctg}^2(\phi_{R1}) = \frac{\mathrm{ctg}^2(\phi_{R0})}{\delta^2 d_0^2} = \frac{q_0}{\delta^2 d_*^2 (1 + p_0)}.
\end{aligned} \tag{6.12}$$

It completes the proof of (6.7). $\square$

Theorem 6.3. *Let the mappings Φ, π and the set Ω are defined as in lemma 6.2. Let us consider matrix $Y_0 \in \pi^{-1}(\Omega)$ and the projector splitting integrator $I(A, D)$ that is defined by* (1). *Then the sequence $Y_k = I(Y_{k-1}, \Phi(Y_{k-1}) - Y_{k-1})$ converges to $Y_* = d_* X_\perp$.*

Proof. We apply Lemma 6.2

$$\pi(Y_k) = f(\pi(Y_{k-1})) = f^{*k}(\pi(Y_0))$$

and then, using Lemma 6.1, we have

$$\lim_{k\to\infty} \pi(Y_k) = \lim_{k\to\infty} f^{*k}(\pi(Y_0)) = \{0, 0\}.$$

Lemma 6.2 guarantees that squared cotangents of left and right angles go to zero, so $\lim\limits_{k\to\infty} Y_k = Y_*$. $\square$

Remark 6.1. *Note that the condition $1 < \delta^2 + \delta^6$ (it requites $\delta > 0.8$) significantly restricts the usage of Theorem 6.3. But our numerical experiments show that the projector splitting method might not converge in computer arithmetics in the case this condition does not hold.*

7. Numerical examples.

7.1. Typical case.

We consider the "linear" contraction mapping

$$\Phi : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}, \quad \Phi(X) = X_* + Q(X - X_*),$$

where X and X^* are rank-r $n \times m$ matrices, Q is a linear operator (on matrices), $n = m = 40$, $r = 7$, $\|Q\| < 0.8$ let us denote singular values of X_* as $\sigma_i, 1 \leq i \leq r$. The typical case corresponds to $\sigma_1/\sigma_r \approx 10$. It shows that the orthogonal part converges quadratically.

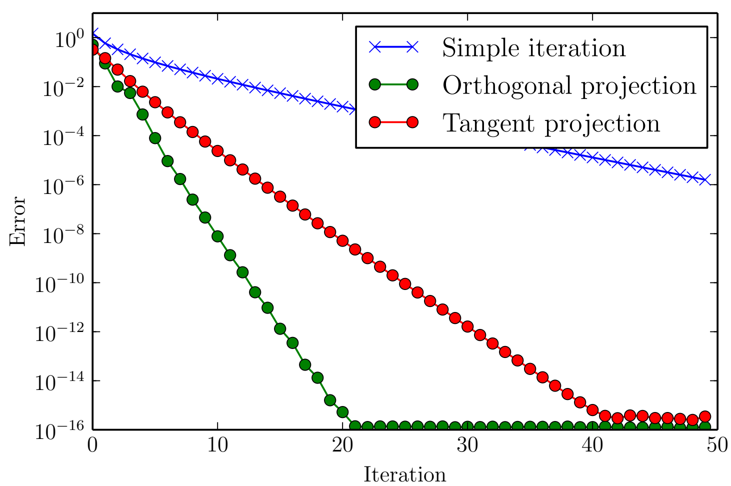


Fig. 7.1: Convergence rates for typical case.

7.2. Stair case.

The stair case corresponds to the same n, m, r and exponentially decaying singular values $\sigma_k = 10^{4-2k}, \quad 1 \leq k \leq 7, \quad \sigma_1/\sigma_r = 10^{12}$. The results are shown on the Figure 7.2. The orthogonal component decays quadratically until the next singular value is achieved. Meanwhile, the tangent component decays linearly, and once it hits the same singular value, the orthogonal component drops again. The steps on the 'stair' correspond to the singular values of X_*.

Numerical experiments show that the projector splitting method has "componentwise" convergence. Until the first j singular components of the current point converge to the first j singular components of the fixed point, the last $r - j$ components of the X_k are "noisy" and do not contain useful information.

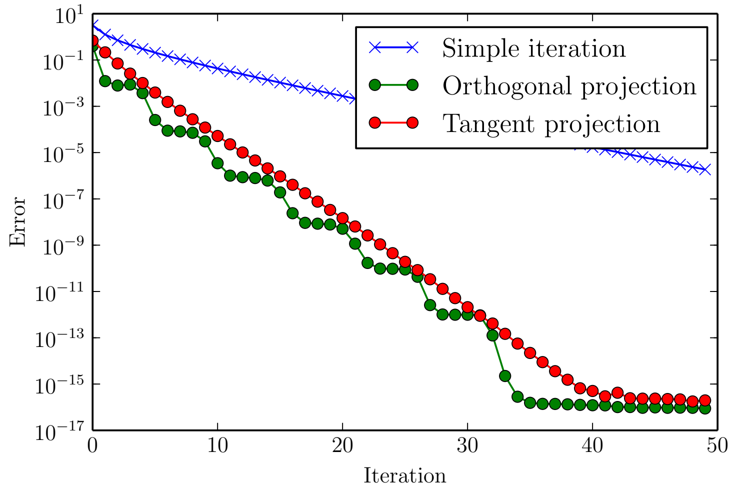


Fig. 7.2: Convergence rates for staircase.

7.3. Counter-example case. For the following experiment we consider "non-linear" contraction mapping

$$\Phi(X) = X_* + \delta \|X - X_*\| X_\perp,$$

where X is a 2×2 matrix, $X_* = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $X_\perp = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $\delta = 0.5$. It shows that original projector splitting method fails and converges to another stationary point. Nevertheless this stationary point is unstable and to show that we introduce a perturbed projector splitting method:

$$Y_{k+1}^{pert} = I(Y_k^{pert}, \Phi(Y_k^{pert}) - Y_k^{pert} + R_k),$$

where R_k is a $n \times m$ matrix with elements taken from the normal distribution $\mathcal{N}\left(0, \frac{1}{100nm} \|\Phi(Y_k^{pert}) - Y_k^{pert}\|\right)$. The convergence is shown at Figure 7.3:

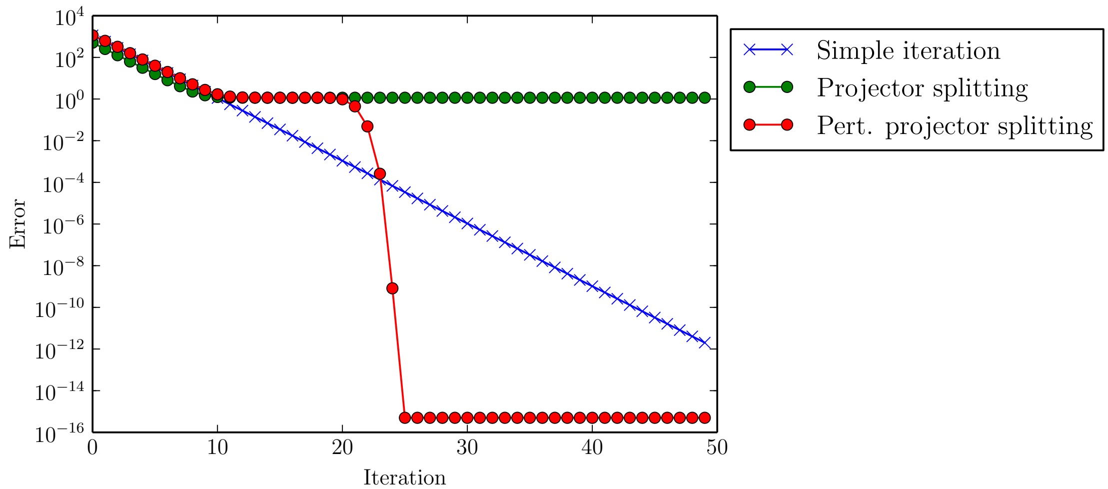


Fig. 7.3: Convergence rates for 'bad functional'

8. Related work. Projector splitting method arises naturally as a numerical integrator for dynamical low-rank approximation of ODE [9, 8] and was originally proposed in [6]. In this paper we focused on the properties of the projector splitting method as the retraction onto low-rank manifold [1]. It was compared with another retraction methods in the survey [2].

Close results about convergence in the presence of small singular values were obtained in [5]. The problem formulation is as follows. Let $X(t)$ be the solution of the ordinary differential equation (ODE):

$$\begin{aligned} \dot{X}(t) = F(t, X(t)), \quad X(0) = X_0, \quad & X(t) \in \mathbb{R}^{n\times m}, && t \in [0, T], \\ \|F(t, X_1) - F(t, X_2)\| \leq L, \quad & \forall X_1, X_2 \in \mathbb{R}^{n\times m}, && \forall t \in [0, T], \\ \|F(t, X)\| \leq B, \quad & \forall X \in \mathbb{R}^{n\times m}, && \forall t \in [0, T]. \end{aligned}$$

We want to obtain approximation to stationary point X_*: $F(t, X_*) = 0$. We seek for low-rank approximation $Y(t)$ to $X(t)$ and $Y(t)$ satisfies the modified ODE:

$$\dot{Y}(t) = P(Y(t))F(t, Y(t)), \quad Y(0) = Y_0, \quad \operatorname{rank} Y(t) = r,$$

where $P(Y(t))$ is a projector onto the subspace determined by $Y(t)$. [5, Theorem 2.1] states that numerical approximation $\widetilde{Y}(t)$ is stable despite the presence of small singular values of $Y(t)$. However, this result cannot be directly applied to optimization problems and F should satisfy certain restrictions.

Another close result is a guaranteed local linear convergence for alternating least squares optimization scheme in convex optimization problems [11]. Also local convergence results are obtained for modified alternating least squares scheme, such as maximum block improvement [7] and alternating minimal energy [3], but for these methods the low-rank manifold changes at every step.

9. Conclusions and perspectives. Our numerical results show that the staircase is a typical case for linear contraction mappings. However, conditions of the proved theorem cover only convergence at the last "step" on the stair. We plan to formulate conditions for the contraction mapping Φ for which "component-wise" convergence as for stair case is guaranteed. Our current hypothesis is that the "extended" mapping $\Phi_m(X, X_*)$ should also satisfy the contraction property for X_*. It will be very interesting to explain the nature of the stair case convergence.

Another important topic for further research is to determine a viable "a-posteriori" error indicator, since we do not know the orthogonal component. This will allow to develop rank-adaptive projector splitting based scheme.

The main conclusion of this paper is that projected iterations are typically as fast as the unprojected ones. We plan to generalize the paper results for tensor case.

Acknowledgements. This work was supported by Russian Science Foundation grant 14-11-00659. We thank Prof. Dr. Christian Lubich and Hanna Walach for fruitful discussions about projector splitting scheme and retractions on a low-rank manifold. We also thank Maxim Rakhuba for his help for improving the manuscript.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.

[2] P.-A. Absil and I. V. Oseledets, *Low-rank retractions: a survey and new results*, Comput. Optim. Appl., 62 (2015), pp. 5–29.

[3] S. V. Dolgov and D. V. Savostyanov, *Alternating minimal energy methods for linear systems in higher dimensions*, SIAM J. Sci. Comput., 36 (2014), pp. A2248–A2271.

[4] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel, *Differential-geometric Newton method for the best rank-(r_1, r_2, r_3) approximation of tensors*, Numerical Algorithms, 51 (2009), pp. 179–194.

[5] E. Kieri, C. Lubich, and H. Walach, *Discretized dynamical low-rank approximation in the presence of small singular values*, to appear in SIAM J. Numer. Anal., (2015).

[6] O. Koch and C. Lubich, *Dynamical low-rank approximation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 434–454.

[7] Z. Li, A. Uschmajew, and S. Zhang, *On convergence of the maximum block improvement method*, SIAM J. Optimiz., 25 (2015), pp. 210–233.

[8] C. Lubich and I. V. Oseledets, *A projector-splitting integrator for dynamical low-rank approximation*, BIT Numer. Math., 54 (2014), pp. 171–188.

[9] C. Lubich, I. V. Oseledets, and B. Vandereycken, *Time integration of tensor trains*, SIAM J. Numer. Anal., 53 (2015), pp. 917–941.

[10] Y. Ma and Y. Fu, *Manifold learning theory and applications*, CRC press, 2011.

[11] T. Rohwedder and A. Uschmajew, *On local convergence of alternating schemes for optimization of convex problems in the tensor train format*, SIAM J. Numer. Anal., 51 (2013), pp. 1134–1162.

[12] H. Sato and T. Iwai, *A new, globally convergent Riemannian conjugate gradient method*, Optimization, 64 (2015), pp. 1011–1031.

[13] C. Udriste, *Convex functions and optimization methods on Riemannian manifolds*, vol. 297, Springer Science & Business Media, 1994.