NONLINEAR POWER-LIKE ITERATION BY POLAR DECOMPOSITION AND ITS APPLICATION TO TENSOR APPROXIMATION

MOODY T. CHU*

Abstract. In order to improve two factor matrices simultaneously for the low rank tensor approximation in the Tucker form, a special system of nonlinear matrix equations over a certain product Stiefold manifold must be resolved at every update. The solution to the system consists of orbit varieties invariant under the orthogonal group action, which thus imposes challenges on it analysis. This paper proposes a scheme similar to the power method for subspace iterations except that the polar decomposition is used as the normalization process and that the iteration applies to both the orbits and cross-sections. The notion of quotient manifold is employed to factor out the effect of orbital solutions. The dynamics of the iteration is completely characterized. An isometric isomorphism between the tangent spaces of two properly identified Riemannian manifolds is established to lend a hand to the proof of convergence,

Key words. Tucker nearest problem, nonlinear matrix equation, quasi-power iteration, polar decomposition, Riemannian manifold, quotient manifold, orbit variety

AMS subject classifications. 65F10, 15A24, 65H10, 15A72, 58D19

1. Introduction.

Let $\mathscr{S}(p,q)$ denote the Stiefel manifold

$$\mathscr{S}(p,q) := \{Q \in \mathbb{R}^{p\times q} | Q^\top Q = \mathrm{I}_q\}. \tag{1.1}$$

Suppose that $\mathcal{A} : \mathscr{S}(p_2,q_2) \times \mathscr{S}(p_2,q_2) \to \mathbb{R}^{p_1\times p_1}$ and $\mathcal{B} : \mathscr{S}(p_1,q_1) \times \mathscr{S}(p_1,q_1) \to \mathbb{R}^{p_2\times p_2}$ are two given bilinear maps with a certain symmetric properties which will be specified in the subsequent discussion. This paper concerns about solving the nonlinear matrix equation

$$\left\{\begin{array}{rcl} \mathcal{A}(Q_2,Q_2)Q_1 &=& Q_1P_1, \\ \mathcal{B}(Q_1,Q_1)Q_2 &=& Q_2P_2, \end{array}\right. \tag{1.2}$$

where the products on the right-hand side of (1.2) are the polar decompositions of the matrices on the left-hand side, respectively. Such a problem arises as a basic computational mechanism when solving the so called Tucker nearest problem. We can rid of the reference to P_1 and P_2 by the substitution of

$$\left\{\begin{array}{rcl} P_1 &=& Q_1^\top \mathcal{A}(Q_2,Q_2)Q_1, \\ P_2 &=& Q_2^\top \mathcal{B}(Q_1,Q_1)Q_2, \end{array}\right. \tag{1.3}$$

provided that the products on the right hand sides are known to be symmetric and positive semi-definite. The system (1.2) is thus cast as an under-determined polynomial system

$$\left\{\begin{array}{rcl} \mathcal{A}(Q_2,Q_2)Q_1 &=& Q_1{Q_1}^\top \mathcal{A}(Q_2,Q_2)Q_1, \\ \mathcal{B}(Q_1,Q_1)Q_2 &=& Q_2{Q_2}^\top \mathcal{B}(Q_1,Q_1)Q_2, \\ {Q_1}^\top Q_1 &=& \mathrm{I}_{q_1}, \\ {Q_2}^\top Q_2 &=& \mathrm{I}_{q_2}, \end{array}\right. \tag{1.4}$$

in the unknowns $Q_1 \in \mathbb{R}^{p_1\times q_1}$ and $Q_2 \in \mathbb{R}^{p_2\times q_2}$, whose solutions will be shown to have rich algebraic properties known as orbit varieties [4].

To motivate where the system of nonlinear matrix equations arises and why it is useful, we briefly outline the background of the Tucker nearest problem. Low rank approximation is one principal tool of great power and interest when dealing with entangled and large-scale data. In recent years, scientists and practitioners

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu) This research was supported in part by the National Science Foundation under grant DMS-1316779.

turn to higher-dimensional arrays, i.e., tensors, for the advantage of greater descriptive flexibility and more fine-grained data collection. Inevitably, the sizes grow rapidly and content analytics becomes a much more challenging task. Low rank tensor approximations, if done properly, are appealing for at least the benefits of storage saving and feature finding. Among a variety of structured or unstructured low rank tensor approximations, the Tucker nearest problem is one of the most fundamental and important formulations with versatile applications.

For convenience, let $[\![k]\!]$ denote the set $\{1, \ldots, k\}$. Let $\mathbf{a} \circ \mathbf{b}$ denote the tensor product enumerated in such a way that, if $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$, then its vectorization is given by

$$\mathbf{vec}(\mathbf{a} \circ \mathbf{b}) = \mathbf{vec}(\mathbf{a}\mathbf{b}^\top) = [a_1 b_1, a_2 b_1, \ldots, a_m b_1, a_1 b_2, \ldots, a_m b_n]^\top. \tag{1.5}$$

Given an order-k tensor $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$ and a fixed rank parameter $\mathbf{r} = (r_1, \ldots, r_k)$ of positive integers, the Tucker nearest problem is to find scalars $c_{j_1,\ldots,j_k}$ and vectors $\mathbf{v}_{j_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell}$, $j_\ell = 1, \ldots, r_\ell$, $\ell = 1, \ldots, k$, such that

$$h(C, V^{(1)}, \ldots, V^{(k)}) := \| \sum_{j_1=1}^{r_1} \cdots \sum_{j_k=1}^{r_k} c_{j_1,\ldots,j_k} \mathbf{v}_{j_1}^{(1)} \circ \ldots \circ \mathbf{v}_{j_k}^{(k)} - T \|_F, \tag{1.6}$$

subject to the condition that

$$V^{(\ell)} := [\mathbf{v}_1^{(\ell)}, \ldots, \mathbf{v}_{r_\ell}^{(\ell)}] \in \mathscr{S}(I_\ell, r_\ell), \quad \ell \in [\![k]\!], \tag{1.7}$$

is minimized. The collections

$$C := [c_{j1,\ldots,j_k}] \in \mathbb{R}^{r_1 \times \ldots \times r_k} \tag{1.8}$$

and $V^{(\ell)}$ are referred to respectively as the core tensor and the factor matrix in the literature. It can be argued that, given $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$, $\ell \in [\![k]\!]$, the optimal C is given by [1, Formula (12)]

$$\mathbf{vec}(C) = ({V^{(k)}}^\top \otimes \ldots \otimes {V^{(1)}}^\top)\mathbf{vec}(T), \tag{1.9}$$

In this way, the Tucker nearest problem is equivalent to the problem of maximizing the Frobenius norm of the core tensor C defined in (1.9) subject to the constraint that $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$, $\ell \in [\![k]\!]$.

A conventional approach to maximizing $\|C\|_F$ is to matricize (1.9) in terms of the mode-d folding. Specifically, it can be shown that [1, Formula (11)]

$$C_{(d)} = {V^{(d)}}^\top \underbrace{T_{(d)}(V^{(k)} \otimes \ldots \otimes V^{(d+1)} \otimes V^{(d-1)} \otimes \ldots \otimes V^{(1)})}_{\Upsilon^{[d]}}, \tag{1.10}$$

where d can be any integer in the set $[\![k]\!]$ and $T_{(d)} \in \mathbb{R}^{I_d \times \prod_{\ell \neq d} I_\ell}$ is a rearrangement of T such that the element $\tau_{i_1,\ldots,i_{d-1},i,i_{d+1},\ldots,i_k}$ of the tensor T is placed at the (i, j) entry of the matrix $T_{(d)}$ with

$$j := i_1 + \sum_{\zeta \neq d} (i_\zeta - 1) \prod_{\eta=1}^{\zeta-1} I_\eta.$$

Such a reformulation motivates the notion of alternating directions that updates one factor matrix $V^{(d)}$ a time for $d \in [\![k]\!]$. That is, for a fixed $\Upsilon^{[d]}$, we solve the optimization problem

$$\max_{V^{(d)} \in \mathscr{S}(I_d, r_d)} f(V^{(d)}) := \frac{1}{2} \| {V^{(d)}}^T \Upsilon^{[d]} \|_F^2 \tag{1.11}$$

for $V^{(d)}$ and repeat this procedure by varying d. It can easily be proved that the optimal $V^{(d)} \in \mathscr{S}(I_d, r_d)$ is given by the first r_d left singular vectors of $\Upsilon^{[d]}$.

In a similar vein, it is possible to reorganize the elements in such a way that we can deal with two updates of the factor matrix simultaneously. See [1, Formula (10)]. In order to accomplish this two-factor update, we need the mechanism for solving the constrained optimization problem

$$\max_{Q_1 \in \mathscr{S}(p_1,q_1), Q_2 \in \mathscr{S}(p_2,q_2)} \quad g(Q_1, Q_2) := \frac{1}{2}\|\Omega(Q_2 \otimes Q_1)\|_F^2, \tag{1.12}$$

where $\Omega \in \mathbb{R}^{m \times p_1 p_2}$ with $m \geq p_1 p_2$ is a fixed matrix. The first order optimality condition of (1.12) results in the nonlinear matrix equation (1.2).

We now give more specifics of the first order optimality condition. Let $\Theta := \Omega^\top \Omega(Q_2 \otimes Q_1) \in \mathbb{R}^{p_1 p_2 \times q_1 q_2}$ be partitioned as a $p_2 \times q_2$ block matrix with blocks of size $p_1 \times q_1$, which is then regarded as an order-4 tensor $\Theta = [\theta_{i_1,i_2,j_1,j_2}]$ in $\mathbb{R}^{p_1 \times q_1 \times p_2 \times q_2}$. For $A = [a_{i_1,i_2}] \in \mathbb{R}^{p_1 \times q_1}$ and $B = [b_{j_1,j_2}] \in \mathbb{R}^{p_2 \times q_2}$, define the multiplications

$$\Theta \circledast_{\{1,2\}} B := \left[\sum_{j_1=1}^{p_2} \sum_{j_2=1}^{q_2} \theta_{i_1,i_2,j_1,j_2} b_{j_1,j_2} \right] \in \mathbb{R}^{p_1 \times q_1} \tag{1.13}$$

and

$$\Theta \circledast_{\{3,4\}} A := \left[\sum_{i_1=1}^{p_1} \sum_{i_2=1}^{q_1} \theta_{i_1,i_2,j_1,j_2} a_{i_1,i_2} \right] \in \mathbb{R}^{p_2 \times q_2}, \tag{1.14}$$

respectively. Then, at a critical point (Q_1, Q_2) of (1.12), the projection of the gradient $\nabla g(Q_1, Q_2)$ onto the product space $\mathscr{S}(p_1, q_1) \times \mathscr{S}(p_2, q_2)$ must be zero. Upon exploiting the topology of the Stiefel manifold, it can be shown that the optimal matrices $Q_1 \in \mathscr{S}(p_1, q_1)$ and $Q_2 \in \mathscr{S}(p_2, q_2)$ for the problem (1.12) are the matrices satisfying

$$\begin{cases} \Theta \circledast_{\{1,2\}} Q_2 &= Q_1 P_1, \\ \Theta \circledast_{\{3,4\}} Q_1 &= Q_2 P_2, \end{cases} \tag{1.15}$$

where the products on the right-hand side of (1.15) stand for the polar decompositions of the matrices on the left-hand side, respectively. We shall further explore the structure of (1.15) to cast it in the form (1.2).

This work is about a numerical procedure and the associated convergence analysis for solving the system (1.15) in particular and the system (1.2) in general. Even though the scope of this paper is limited to only a special type of nonlinear matrix equation, solving (1.15) is an indispensable part for solving the Tucker nearest problem as outlined above. To our knowledge, current updating techniques for the Tucker nearest problem have been limited to one factor matrix a time. The lack of simultaneous factor matrix updating might be attributable to the nonlinearity such as that involved in (1.15) that seems too complicated to handle. Our contribution in this paper therefore is innovative. Additionally, the equation (1.2) resembles a nonlinear eigenvalue problem with eigenvalues being replaced by positive semi-definite matrices. Our numerical procedure resembles the conventional simultaneous subspace iteration where the normalization is being carried out by positive semi-definite matrices. In all, there might be enough mathematics of interest in this study.

2. Basics. The tensor Θ involved in (1.15) is a function of both unknown matrices Q_1 and Q_2. So the polar decompositions on the right-hand sides are implicitly defined. Though it is true that the symmetric and positive semi-definite matrices P_1 and P_2 are uniquely determined from

$$\begin{cases} P_1 &= ((\Theta \circledast_{\{1,2\}} Q_2)^\top (\Theta \circledast_{\{1,2\}} Q_2))^{\frac{1}{2}}, \\ P_2 &= ((\Theta \circledast_{\{3,4\}} Q_1)^\top (\Theta \circledast_{\{3,4\}} Q_1))^{\frac{1}{2}}, \end{cases} \tag{2.1}$$

any attempt of substituting them into (1.15) only makes the nonlinearity more tangled. We have to search for some indirect approach.

2.1. Bilinear formulation. The operator-like multiplications defined by (1.13) and (1.14) and the way the optimality condition (1.15) comes into sight are natural from tensor point of view [10, 11], but appear cumbersome to manipulate. The following reformulation sheds better insight into the symmetry innate to the system (1.15), which will help motivate a way to solve the polar equations.

LEMMA 2.1. *Suppose $Q_2 = [\psi_{i,j}], \widetilde{Q}_2 = [\widetilde{\psi}_{i,j}] \in \mathbb{R}^{p_2 \times q_2}$. Partition the matrix $\Omega^\top \Omega \in \mathbb{R}^{p_1 p_2 \times p_1 p_2}$ into blocks $\Omega^\top \Omega = [W_{ij}]$ with $W_{i,j} \in \mathbb{R}^{p_1 \times p_1}$ with $i, j \in [\![p_2]\!]$. Then the multiplication (1.13) can be expressed as*

$$\Theta \circledast_{\{1,2\}} \widetilde{Q}_2 = \mathcal{A}(Q_2, \widetilde{Q}_2) Q_1, \tag{2.2}$$

where $\mathcal{A}(Q_2, \widetilde{Q}_2)$ is a matrix in $\mathbb{R}^{p_1 \times p_1}$ defined by

$$\mathcal{A}(Q_2, \widetilde{Q}_2) := \sum_{i=1}^{p_2} \sum_{j=1}^{q_2} \widetilde{\psi}_{i,j} (\sum_{k=1}^{p_2} W_{i,k} \psi_{k,j}). \tag{2.3}$$

Proof. Let the colon ":" denote unspecified arrays of indices. The (i, j)-th block of Θ is a $p_1 \times q_1$ matrix given by

$$\theta_{:,:,i,j} = (\sum_{k=1}^{p_2} W_{i,k} \psi_{k,j}) Q_1.$$

By definition (1.13),

$$\Theta \circledast_{\{1,2\}} \widetilde{Q}_2 = \sum_{i=1}^{p_2} \sum_{j=1}^{q_2} \theta_{:,:,i,j} \tilde{\psi}_{i,j}.$$

The relationship (2.2) follows by factoring Q_1 out of the summation. □

A similar relationship holds for the multiplication (1.14).

LEMMA 2.2. *Suppose $Q_1, \widetilde{Q} \in \mathbb{R}^{p_1 \times q_1}$ and $\Omega^\top \Omega = [W_{ij}]$ is partitioned in the same way as in Lemma 2.1. Then*

$$\Theta \circledast_{\{3,4\}} \widetilde{Q}_1 = \mathcal{B}(Q_1, \widetilde{Q}_1) Q_2, \tag{2.4}$$

where $\mathcal{B}(Q_1, \widetilde{Q}_1) \in \mathbb{R}^{p_2 \times p_2}$ with entries defined by

$$\mathcal{B}(Q_1, \widetilde{Q}_1) := \left[\langle \widetilde{Q}_1, W_{i,j} Q_1 \rangle \right]. \tag{2.5}$$

Note that $\mathcal{A}(Q_2, \widetilde{Q}_2) \in \mathbb{R}^{p_1 \times p_1}$ is bilinear in Q_2 and $\widetilde{Q}_2$ and $\mathcal{B}(Q_1, \widetilde{Q}_1) \in \mathbb{R}^{p_2 \times p_2}$ is bilinear in Q_1 and $\widetilde{Q}_1$. In particular, the first optimality conditions (1.15) is now expressed as finding $Q_1 \in \mathscr{S}(p_1, q_1)$ and $Q_2 \in \mathscr{S}(p_2, q_2)$ such that (1.2) with the specially defined $\mathcal{A}$ and $\mathcal{B}$ is satisfied. We think that (1.2) is in a much manageable form than (1.15). The remaining task will be on developing a numerical method for solving (1.2) and proving its convergence.

2.2. Symmetry and invariance. The operators $\mathcal{A}$ and $\mathcal{B}$ defined by (2.3) and (2.5) enjoy two important properties[1] — a sense of symmetry as well as the invariance under the right group action by orthogonal matrices — which we characterize below. These properties are important tools for our convergence analysis.

[1]These two characteristics mentioned are innate to our operators $\mathcal{A}$ and $\mathcal{B}$ specifically derived for tensor applications. For general (1.2), if we assume that the bilinear operators $\mathcal{A}$ and $\mathcal{B}$ satisfy the symmetry (2.6), (2.7), and the invariance (2.8). then the remaining discussion can be equally applied.

LEMMA 2.3. *For any $Q_1, \widetilde{Q}_1 \in \mathbb{R}^{p_1 \times q_1}$ and $Q_2, \widetilde{Q}_2 \in \mathbb{R}^{p_2 \times q_2}$, there is a symmetry in the sense that*

$$\begin{cases} \mathcal{A}(Q_2, \widetilde{Q}_2) &= \mathcal{A}(\widetilde{Q}_2, Q_2)^\top, \\ \mathcal{B}(Q_1, \widetilde{Q}_1) &= \mathcal{B}(\widetilde{Q}_1, Q_1)^\top. \end{cases} \tag{2.6}$$

It also holds that

$$\langle \mathcal{B}(Q_1, \widetilde{Q}_1)Q_2, \widetilde{Q}_2\rangle = \langle \mathcal{A}(Q_2, \widetilde{Q}_2)Q_1, \widetilde{Q}_1\rangle. \tag{2.7}$$

Proof. Sine W is symmetric, $W_{i,k}^\top = W_{k,i}$ for all $i, k \in [\![p_2]\!]$. The symmetry in (2.6) is obvious from the definitions (2.3) and (2.5). We also see that

$$\begin{aligned} \langle \mathcal{B}(Q_1, \widetilde{Q}_1)Q_2, \widetilde{Q}_2\rangle &= \sum_{i=1}^{p_2}\sum_{j=1}^{q_2} \widetilde{\psi}_{ij}(\sum_{k=1}^{p_2}\langle \widetilde{Q}_1, W_{ik}Q_1\rangle \psi_{kj}) \\ &= \langle \widetilde{Q}_1, \sum_{i=1}^{p_2}\sum_{j=1}^{q_2} \widetilde{\psi}_{ij}(\sum_{k=1}^{p_2} W_{ik}\psi_{kj})Q_1\rangle = \langle \widetilde{Q}_1, \mathcal{A}(Q_2, \widetilde{Q}_2)Q_1\rangle \end{aligned}$$

which offers another kind of symmetry. □

COROLLARY 2.4. *If the bilinear maps $\mathcal{A}$ and $\mathcal{B}$ are defined via (2.3) and (2.5), respectively, then both $\mathcal{A}(Q_2, Q_2) \in \mathbb{R}^{p_1 \times p_1}$ and $\mathcal{B}(Q_1, Q_1) \in \mathbb{R}^{p_2 \times p_2}$ are symmetric and positive semi-definite matrices.*

Let the special case $\mathscr{S}(q, q)$ of orthogonal matrices be denoted as $\mathscr{O}(q)$. It should be straightforward to see from the definitions of (2.3) and (2.5) that the bilinear maps $\mathcal{A}$ and $\mathcal{B}$ are invariant under the right group action by orthogonal matrices in the following sense.

LEMMA 2.5. *If $U_1 \in \mathscr{O}(q_1)$ and $U_2 \in \mathscr{O}(q_2)$ are orthogonal, then*

$$\begin{cases} \mathcal{A}(Q_2U_2, \widetilde{Q}_2U_2) &= \mathcal{A}(Q_2, \widetilde{Q}_2), \\ \mathcal{B}(Q_1U_1, \widetilde{Q}_1U_1) &= \mathcal{B}(Q_1, \widetilde{Q}_1). \end{cases} \tag{2.8}$$

LEMMA 2.6. *If (Q_1, Q_2) is a solution to (1.2), then the entire orbit*

$$\mathscr{X}(Q_1, Q_2) := \{(Q_1U_1, Q_2U_2) | U_1 \in \mathscr{O}(q_1), U_2 \in \mathscr{O}(q_2)\} \tag{2.9}$$

by the right group action is also a solution. The solutions to the system (1.2) therefore are not isolated.

Proof. If (Q_1, Q_2) is a solution, then by (2.8)

$$\begin{cases} \mathcal{A}(Q_2U_2, Q_2U_2)Q_1U_1 &= Q_1U_1(U_1^\top P_1U_1), \\ \mathcal{B}(Q_1U_1, Q_1U_1)Q_2U_2 &= Q_2U_2(U_2^\top P_2U_2), \end{cases} \tag{2.10}$$

implying that (Q_1U_1, Q_2U_2) for any $U_1 \in \mathscr{O}(q_1)$ and $U_2 \in \mathscr{O}(q_2)$ also satisfies (1.2). □

For our application (1.12), note that every such a right actioned pair (Q_1U_1, Q_2U_2) results in the same objective value as $g(Q_1, Q_2)$ because $(Q_2U_2) \otimes (Q_1U_1) = (Q_2 \otimes Q_1) \otimes (U_1 \otimes U_2)$, whereas $U_1 \otimes U_2$ is itself orthogonal [7].

2.3. Quotient manifold. By Corollary 2.4, the constrained system (1.2) is equivalent to the free system (1.4). The latter is independent of P_1 and P_2. A careful check reveals that the first subsystem in (1.4) involves only $(p_1 - q_1)q_1$ independent equations while third subsystem involves $\frac{q_1(q_1+1)}{2}$ equations. Similar counts hold for the second and the fourth subsystems. In total, the system (1.4) is under-determined and has $\frac{q_1(q_1-1)}{2} +$

$\frac{q_2(q_2-1)}{2}$ degrees of freedom which are precisely the dimensionality of $\mathscr{O}(q_1) \times \mathscr{O}(q_2)$. Each orbit $\mathscr{X}$ is isomorphic to $\mathscr{O}(q_1) \times \mathscr{O}(q_2)$, but there might be disjoint orbits as the following example shows.

Example 1. Consider the case $(Q_1, Q_2) \in \mathscr{S}(3,2) \times \mathscr{S}(2,1)$. Let $Z \in \mathbb{R}^{3\times 3}$ be an arbitrary symmetric and positive definite matrix such that its eigenvalues are distinct and that its smallest eigenvalue is sufficiently away from zero. Define

$$W = \begin{bmatrix} Z & 0_3 \\ 0_3 & \mathrm{I}_3 \end{bmatrix}$$

which is considered as an order-4 tensor in $\mathbb{R}^{3\times 3\times 2\times 2}$. Take $Q_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then

$$\mathcal{A}(Q_2, Q_2) = Z.$$

Take Q_1 to be any two of the three orthonormal eigenvectors of Z. Then

$$\mathscr{B}(Q_1, Q_1) = \begin{bmatrix} \mathrm{trace}(Q_1^\top Z Q_1) & 0 \\ 0 & 2 \end{bmatrix}$$

which has Q_1 as its eigenvector. It is readily observable that (1.4) is satisfied, but there are at least three disjoint orbits because the eigenvector of Z are linearly independent.

Given that the system (1.4) has orbit varieties and every element in the orbit results in the same objective value, it is sufficient to reconsider the system as over the quotient manifold[2]

$$\mathscr{S}(p_1, q_1)/\mathscr{O}(q_1) \times \mathscr{S}(p_2, q_2)/\mathscr{O}(q_2). \tag{2.11}$$

In this way, we "shrink" an orbit of infinitely many points in $\mathscr{S}(p_1, q_1) \times \mathscr{S}(p_2, q_2)$ to a single point in the quotient manifold over which the polynomial system (1.4) has the same numbers of equations and unknowns.

Without causing ambiguity, we shall use the same notation Q_i, $i = 1, 2$, to represent interchangeably between the orbit

$$[Q_i] := \{Q_i U | U \in \mathscr{O}(q_i)\} \in \mathscr{S}(p_i, q_i)/\mathscr{O}(q_i) \tag{2.12}$$

and the element $Q_i \in \mathscr{S}(p_i, q_i)$. In what follows, the calculation is applied to points in $\mathscr{S}(p_1, q_1) \times \mathscr{S}(p_2, q_2)$ as usual, but we shall use the induced metric in the quotient manifold $\mathscr{S}(p_1, q_1)/\mathscr{O}(q_1) \times \mathscr{S}(p_2, q_2)/\mathscr{O}(q_2)$ to argue the convergence.

3. Numerical Method. At first glance, the system (1.2) resembles a nonlinear eigenvalue problem [8], except that in place of eigenvalues are the symmetric and positive semi-definite matrices P_1 and P_2. It is natural to formulate an iterative scheme analogous to the conventional power method for eigenvalue computation [5]. To deal with the nonlinearity, the simplest approach might be to alternate directions in the iteration. We thus propose the scheme

$$\begin{cases} [Q_{1,[s+1]}, P_{1,[s+1]}] &= \mathsf{poldec}(\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}), \\ [Q_{2,[s+1]}, P_{2,[s+1]}] &= \mathsf{poldec}(\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}), \end{cases} \tag{3.1}$$

where poldec denote any algorithm for computing the polar decomposition, such as that [6]. We remark again that the expression (3.1) is invariant under the right group action by orthogonal matrices (see (2.10)), so it can be interpreted as an iteration on orbits. Since there is a one-to-one correspondence of elements among orbits,

[2]Though it is not directly relevant to our discussion, in topology it can be proved that since $\mathscr{O}(q)$ is compact, the quotient space $\mathscr{S}(p, q)/\mathscr{O}(q)$ is Hausdorff. Furthermore, since the right group action is free, the quotient space is indeed a manifold.

it suffices to consider the evolution of just one "cross-section" as indicated by the scheme. The notion of cross-sections will be explained in Section 3.3 in the context of Riemannian geometry. The primary concern is what the iterative dynamics will lead to. Our goal is to prove that the sequence $\{(Q_{1,[s]}, Q_{2,[s]})\}$ will converge to a limit point in $\mathscr{S}(p_1, q_1) \times \mathscr{S}(p_2, q_2)$ that solves (1.2).

Recall that if $Z = QP$ is the polar decomposition of $Z \in \mathbb{R}^{p\times q}$, then

$$Q = \underset{U\in\mathscr{S}(p,q)}{\arg\min} \|Z - U\|_F = \underset{U\in\mathscr{S}(p,q)}{\arg\max} \langle Z, U\rangle, \tag{3.2}$$

whereas

$$\langle Z, Q\rangle = \langle P, I\rangle = \mathrm{trace}(P). \tag{3.3}$$

Therefore, the definitions of $Q_{1,[s+1]}$ and $Q_{2,[s+1]}$ in (3.1) enjoy the variational properties of maximizing the traces of $P_{1,[s+1]}$ and $P_{2,[s+1]}$, respectively. These properties will exploited to prove the convergence. We shall argue the convergence in two aspects. We first prove the convergence of the traces in general, which then will be employed to argue the convergence of the iterates under mild assumptions.

3.1. Convergence of traces. Assuming the generic condition that the given Ω is of full column rank, the matrix $W = \Omega^\top\Omega$ is symmetric and positive definite. By (2.3) and (2.5), both $\mathcal{A}(Q_{2,[s]}, Q_{2,[s]}) \in \mathbb{R}^{p_1\times p_1}$ and $\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]}) \in \mathbb{R}^{p_2\times p_2}$ are symmetric and positive semi-definite for all s. Given the algebraic structure in (2.3) and (2.5), the matrices $Q_{1,[s+1]}$ or $Q_{2,[s]}$ that cause $\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})$ or $\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})$ singular must satisfy some very special polynomials whose algebraic varieties are of measure zero. To rigorously describe these varieties might be too distracting for the presentation of this paper, so we will assume that our iterates do not hit these zero-measure sets.

Under such an assumption, we may thus write $Q_{1,[s+1]}$ and $Q_{2,[s+1]}$ explicitly as

$$\begin{cases} Q_{1,[s+1]} &= \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}(Q_{1,[s]}^\top\mathcal{A}^2(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]})^{-\frac{1}{2}} \\ Q_{2,[s+1]} &= \mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}(Q_{2,[s]}^\top\mathcal{B}^2(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]})^{-\frac{1}{2}}. \end{cases} \tag{3.4}$$

We first establish a variational relationship in terms of traces.

LEMMA 3.1. *The sequence $\{(Q_{1,[s]}, Q_{2,[s]})\} \subset \mathscr{S}(p_1, q_1) \times \mathscr{S}(p_2, q_2)$ generated by the scheme (3.1) satisfies the inequalities*

$$\begin{aligned} &\langle\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}, Q_{1,[s+1]}\rangle \le \langle\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}, Q_{2,[s+1]}\rangle \\ \le\ &\langle\mathcal{A}(Q_{2,[s+1]}, Q_{2,[s+1]})Q_{1,[s+1]}, Q_{1,[s+2]}\rangle \le \langle\mathcal{B}(Q_{1,[s+2]}, Q_{1,[s+2]})Q_{2,[s+1]}, Q_{2,[s+2]}\rangle, \end{aligned} \tag{3.5}$$

and equivalently

$$\mathrm{trace}(P_{1,[s+1]}) \le \mathrm{trace}(P_{2,[s+1]}) \le \mathrm{trace}(P_{1,[s+2]}) \le \mathrm{trace}(P_{2,[s+2]}). \tag{3.6}$$

Proof. We claim first that

$$\langle\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}, Q_{1,[s+1]}\rangle \le \langle\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s+1]}, Q_{1,[s+1]}\rangle, \tag{3.7}$$

$$\langle\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}, Q_{2,[s+1]}\rangle \le \langle\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s+1]}, Q_{2,[s+1]}\rangle. \tag{3.8}$$

Using (3.7) and the definition of $Q_{2,[s+1]}$, the first inequality in (3.5) follows from the fact that

$$\begin{aligned} \langle\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s+1]}, Q_{1,[s+1]}\rangle &= \langle\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}, Q_{2,[s]}\rangle \\ &\le \langle\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s]}, Q_{2,[s+1]}\rangle. \end{aligned} \tag{3.9}$$

Likewise, using (3.8), the second inequality in (3.5) follows from

$$\begin{aligned}\langle \mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})Q_{2,[s+1]}, Q_{2,[s+1]}\rangle &= \langle \mathcal{A}(Q_{2,[s+1]}, Q_{2,[s+1]})Q_{1,[s+1]}, Q_{1,[s+1]}\rangle \\ &\le \langle \mathcal{A}(Q_{2,[s+1]}, Q_{2,[s+1]})Q_{1,[s+1]}, Q_{1,[s+2]}\rangle. \end{aligned} \tag{3.10}$$

The third inequality is just a shift of index from the first inequality.

It only remains to prove the claims (3.7) and (3.8). We shall prove only (3.7), as the argument for (3.8) is similar. Write

$$\Delta Q_{1,[s]} := Q_{1,[s+1]} - Q_{1,[s]}. \tag{3.11}$$

Observe that

$$\begin{aligned}&\langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]}, Q_{1,[s+1]}\rangle = \langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]}, Q_{1,[s]} + \Delta Q_{1,[s]}\rangle \\ &= \langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]}, Q_{1,[s]}\rangle + \langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]}, \Delta Q_{1,[s]}\rangle.\end{aligned}$$

The second term is nonnegative because $\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})$ is symmetric and positive semi-definite. By using (3.4), we now write

$$Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]} = (Q_{1,[s]}^\top \mathcal{A}^2(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]})^{\frac{1}{2}} - Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}.$$

We claim that

$$(Q_{1,[s]}^\top \mathcal{A}^2(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]})^{\frac{1}{2}} \succeq Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]} \tag{3.12}$$

in the sense of the the Loewner partial order and thus the matrix $Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})\Delta Q_{1,[s]}$ is positive semi-definite. To see this, observer that

$$\begin{aligned}&Q_{1,[s]}^\top \mathcal{A}^2(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]} - (Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]})^2 \\ &= Q_{1,[s]}^\top \mathcal{A}(Q_{2,[s]}, Q_{2,[s]})(I - Q_{1,[s]}Q_{1,[s]}^\top)\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})Q_{1,[s]}.\end{aligned}$$

Since $Q_{1,[s]} \in \mathscr{S}(p_1, q_1)$, the matrix $I - Q_{1,[s]}Q_{1,[s]}^\top$ is itself symmetric and positive semi-definite. Recall that the square roots of partially ordered symmetric and semi-definite matrices preserve the ordering. We have thus proved all claims. □

Because $\{Q_{1,[s]}\} \subset \mathscr{S}(p_1, q_1)$ and $\{Q_{2,[s]})\} \subset \mathscr{S}(p_2, q_2)$, we see that both sequence $\{\mathrm{trace}(P_{1,[s]})\}$ and $\{\mathrm{trace}(P_{2,[s]})\}$ are bounded. By the monotonicity proved in (3.6), the sequences $\{\mathrm{trace}(P_{1,[s]})\}$ and $\{\mathrm{trace}(P_{2,[s]})\}$ must converge.

3.2. Diminishing increments. We first prove the diminishing increment between successive iterates. Define

$$h(Q_1, \widetilde{Q}_1; Q_2, \widetilde{Q}_2) := \langle \mathcal{B}(Q_1, \widetilde{Q}_1)Q_2, \widetilde{Q}_2\rangle = \langle \mathcal{A}(Q_2, \widetilde{Q}_2)Q_1, \widetilde{Q}_1\rangle. \tag{3.13}$$

It might be informative to summarize the above proof as a sequence of telescoping inequalities

$$\begin{array}{rlrl} & h(Q_{1,[s]}, Q_{1,[s]}; Q_{2,[s]}, Q_{2,[s]}) & \le & h(Q_{1,[s]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s]}) \\ \le & h(Q_{1,[s+1]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s]}) & \le & h(Q_{1,[s+1]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s+1]}) \\ \le & h(Q_{1,[s+1]}, Q_{1,[s+1]}; Q_{2,[s+1]}, Q_{2,[s+1]}) & \le & h(Q_{1,[s+1]}, Q_{1,[s+2]}; Q_{2,[s+1]}, Q_{2,[s+1]}). \end{array} \tag{3.14}$$

Observe that in the process the variables are updated one a time and push up the values of h per step. Observe also the two ends of this chain ensure that the telescoping behavior repeats for all s.

THEOREM 3.2. *Assume that $\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})$ and $\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})$ are positive definite for all but finitely many s. Then $\Delta Q_{1,[s]}$ and $\Delta Q_{2,[s]}$ converge to zero.*

Proof. The chain (3.14) allows us establish the differences

$$h(Q_{1,[s]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s]}) - h(Q_{1,[s]}, Q_{1,[s]}; Q_{2,[s]}, Q_{2,[s]}) = \langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]}) Q_{1,[s]}, \Delta Q_{1,[s]}, \rangle$$
$$h(Q_{1,[s+1]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s]}) - h(Q_{1,[s]}, Q_{1,[s+1]}; Q_{2,[s]}, Q_{2,[s]}) = \langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]}) Q_{1,[s+1]}, \Delta Q_{1,[s]} \rangle.$$

The interlacing of the chain, together with the convergence of traces, ascertains that these differences converge to zero simultaneously. Taking the difference of the above two equations again, we see that

$$\langle \mathcal{A}(Q_{2,[s]}, Q_{2,[s]}) \Delta Q_{1,[s]}, \Delta Q_{1,[s]} \rangle \to 0. \tag{3.15}$$

A similar argument ensures that

$$\langle \mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]}) \Delta Q_{2,[s]}, \Delta Q_{2,[s]} \rangle \to 0. \tag{3.16}$$

Under the assumption that $\mathcal{A}(Q_{2,[s]}, Q_{2,[s]})$ and $\mathcal{B}(Q_{1,[s+1]}, Q_{1,[s+1]})$ are positive definite, the assertion is proved. □

3.3. Convergence of orbits. The Stiefel manifold $\mathscr{S}(p,q)$ is a Riemannian manifold. Its tangent space $\mathfrak{T}_Q\mathscr{S}(p,q)$ at $Q \in \mathscr{S}(p,q)$ is made of matrices $H \in \mathbb{R}^{p\times q}$ in the form [2]

$$H = QK + (\mathrm{I}_p - QQ^\top)W, \tag{3.17}$$

where $K \in \mathbb{R}^{q\times q}$ and $W \in \mathbb{R}^{p\times q}$ are arbitrary, and K is skew-symmetric. The group action by $\mathscr{O}(q)$ on $\mathscr{S}(p,q)$ is proper and free. Thus the set $\mathscr{S}(p,q)/\mathscr{O}(q)$ is a differentiable manifold. The quotient map

$$\pi : \mathscr{S}(p,q) \to \mathscr{S}(p,q)/\mathscr{O}(q) \tag{3.18}$$

is a submersion, that is, its differential

$$d\pi_Q : \mathfrak{T}_Q\mathscr{S}(p,q) \to \mathfrak{T}_{[Q]}(\mathscr{S}(p,q)/\mathscr{O}(q)) \tag{3.19}$$

is surjective for every $Q \in \mathscr{S}(p,q)$. On the other hand, the orbit $[Q]$ is itself a submanifold in $\mathscr{S}(p,q)$ with

$$\mathfrak{T}_Q[Q] = \{QK \in \mathbb{R}^{p\times q} | K \in \mathbb{R}^{q\times q} \text{ is skew-symmetric}\} \subset \mathfrak{T}_Q\mathscr{S}(p,q). \tag{3.20}$$

The null space of $d\pi_Q$ is precisely $\mathfrak{T}_Q[Q]$. It follows that $d\pi_Q$ induces an isomorphism, denoted by $d\pi_Q^*$,

$$d\pi_Q^* : \mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q] \to \mathfrak{T}_{[Q]}(\mathscr{S}(p,q)/\mathscr{O}(q)). \tag{3.21}$$

We thus can identify each element of $\mathfrak{T}_{[Q]}(\mathscr{S}(p,q)/\mathscr{O}(q))$ with an element of $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$. Furthermore, because the group action by $\mathscr{S}(q)$ is isometric, we can impose an inner product on $\mathfrak{T}_{[Q]}(\mathscr{S}(p,q)/\mathscr{O}(q))$ in the same way as an inner product on $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$ via the inverse map ${d\pi_Q^*}^{-1}$.

We now specify the inner product over the quotient space $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$. The notation

$$[\![X]\!] := \{X + V | V \in \mathfrak{T}_Q[Q]\} \tag{3.22}$$

of an orbit in the quotient space $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$ is not to be confused with the orbit $[Q]$ in the quotient manifold $\mathscr{S}(p,q)/\mathscr{O}(q)$. By the decomposition

$$\mathfrak{T}_Q\mathscr{S}(p,q) = \mathfrak{T}_Q[Q] \oplus (\mathfrak{T}_Q[Q])^\perp,$$

we quickly see from (3.17) that if $Q_\perp$ denotes the matrix in $\mathscr{S}(p,p-q)$ such that the augmented matrix $[Q, Q_\perp]$ is in $\mathscr{O}(p)$, then

$$(\mathfrak{T}_Q[Q])^\perp = \{Q_\perp \Psi | \Psi \in \mathbb{R}^{(p-q)\times q} \text{ is arbitrary}\}. \tag{3.23}$$

Any element $X \in \mathfrak{T}_Q\mathscr{S}(p,q)$ has a unique decomposition

$$X = QK_X + Q_\perp \Psi_X, \tag{3.24}$$

where $QK_X \in \mathfrak{T}_Q[Q]$ and $Q_\perp\Psi_X \in (\mathfrak{T}_Q[Q])^\perp$. For any $[\![X]\!], [\![Y]\!] \in \mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$, define

$$\langle [\![X]\!], [\![Y]\!] \rangle := \langle \Psi_X, \Psi_Y \rangle, \tag{3.25}$$

where, without causing ambiguity, we use the same inner product notation $\langle \cdot, \cdot \rangle$ on both sides.

LEMMA 3.3. *The map $\langle [\![X]\!], [\![Y]\!] \rangle$ defined in (3.25) is independent of its representatives and is an inner product over the quotient space $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$.*

Proof. Suppose that $X_1, X_2 \in [\![X]\!]$ and $Y_1, Y_2 \in [\![Y]\!]$. Then

$$\langle [\![X_1]\!], [\![Y_1]\!] \rangle - \langle [\![X_2]\!], [\![Y_2]\!] \rangle = \langle [\![X_1]\!] - [\![X_2]\!], [\![Y_1]\!] \rangle + \langle [\![X_2]\!], [\![Y_1]\!] - [\![Y_2]\!] \rangle.$$

Since $X_1 - X_2 \in \mathfrak{T}_Q[Q]$, it must be that $\Psi_{X_1} = \Psi_{X_2}$. It follows that

$$\langle [\![X_1]\!] - [\![X_2]\!], [\![Y_1]\!] \rangle = \langle \Psi_{X_1} - \Psi_{X_2}, \Psi_{Y_1} \rangle = 0.$$

Likewise, since $\Psi_{Y_1} = \Psi_{Y_2}$, the second term is also zero.

The conjugate symmetry and linearity of (3.25) are obvious. We also check that if $\langle [\![X]\!], [\![X]\!] \rangle = 0$, then $\Psi_X = 0$, implying that $X \in \mathfrak{T}_Q[Q]$. □

COROLLARY 3.4. *The quotient manifold $\mathscr{S}(p,q)/\mathscr{O}(q)$ is a Riemannia manifold equipped with an inner product that is isometric to the inner product in $\mathfrak{T}_Q\mathscr{S}(p,q)/\mathfrak{T}_Q[Q]$ defined by (3.25). Indeed, its tangent space $\mathfrak{T}_{[Q]}(\mathscr{S}(p,q)/\mathscr{O}(q))$ can be identified with the subspace $(\mathfrak{T}_Q[Q])^\perp$.*

It thus follows that $\mathscr{S}(p,q)/\mathscr{O}(q)$ can be endowed with a metric, known as the geodesics, for measuring the distance between orbits. That is,

$$d([Q],[\tilde{Q}]) = \inf_{[\gamma]} \{ \int_0^1 \|[\gamma(t)]'\| dt \}, \tag{3.26}$$

where $[\gamma(t)]$ is a continuous differentiable curve on the manifold $\mathscr{S}(p,q)/\mathscr{O}(q)$ joining $[Q]$ and $[\tilde{Q}]$, $[\gamma(t)]'$ is the corresponding tangent vector in $\mathfrak{T}_{[\gamma(t)]}(\mathscr{S}(p,q)/\mathscr{O}(q))$, and $\|\cdot\|$ is the norm induced by the underlying inner product.

Finally, we are ready to apply the above concept to our algorithm (3.1).

THEOREM 3.5. *Let $\{(Q_{1,[s]}, Q_{2,[s])}\}$ be the sequence generated by the scheme (3.1). Let d_i, $i = 1, 2$, be the metrics defined by (3.26) on $\mathscr{S}(p_i,q_i)/\mathscr{O}_i$, respectively. Then $d_i([Q_{i,[s+1]}], [Q_{i,[s]}])$ converges to zero.*

Proof. To measure the distance between $[Q_{i,[s+1]}]$ and $[Q_{i,[s]}]$, it suffices to consider by the isometric isomorphism described above the pullback curve $\gamma_i(t)$ on $\mathscr{S}(p_i,q_i)$ connecting $Q_{i,[s+1]}$ and $Q_{i,[s]}$ with the identities

$$\|[\gamma_i(t)]'\| = \|d\pi^{-1}_{\gamma_i(t)}[\gamma_i(t)]'\| = \|[\![\gamma_i'(t)]\!]\| = \|\gamma_i(t)_\perp \Psi_{\gamma_i'(t)}\| = \|\Psi_{\gamma_i'(t)}\|. \tag{3.27}$$

On one hand, note that by Lemma 3.3 the metrics d_i are independent of the representatives. On the other hand, note that by Theorem 3.2 we have already shown that in one "cross-section", namely, the sequence $\{(Q_{1,[s]}, Q_{2,[s])}\}$ generated by (3.1), the increments $\Delta Q_{1,[s]}$ and $\Delta Q_{2,[s]}$ converge to zero. It follows by continuity that $d_i([Q_{i,[s+1]}], [Q_{i,[s]}])$ converges to zero for $i = 1, 2$. □

Recall the following lemma from the theory of parameter continuation [9, Theorem 7.1.1] concerns geometrically isolated solutions to a generic polynomial system.

LEMMA 3.6. *Let $P(\mathbf{z};\mathbf{q})$ be a system of n polynomials in variables $\mathbf{z} \in \mathbb{C}^n$ and parameters $\mathbf{q} \in \mathbb{C}^m$. Let $\mathcal{N}(\mathbf{q})$ denote the number of geometrically isolated solutions to $P(\mathbf{z};\mathbf{q}) = 0$ over the algebraically closed complex space. Then,*

1. *$\mathcal{N}(\mathbf{q})$ is finite, and it is the same, say $\mathcal{N}$, for almost all $\mathbf{q} \in \mathbb{C}^m$;*
2. *For all $\mathbf{q} \in \mathbb{C}^m$, $\mathcal{N}(\mathbf{q}) \leq \mathcal{N}$;*
3. *The subset of $\mathbb{C}^m$ where $\mathcal{N}(\mathbf{q}) = \mathcal{N}$ is a Zariski open set. That is, the exceptional subset of $\mathbf{q} \in \mathbb{C}^m$ where $\mathcal{N}(\mathbf{q}) < \mathcal{N}$ is an affine algebraic set contained within an algebraic set of codimension one.*

Since $\mathbb{R}^n$ (indeed, the closure of any infinite subset) is Zariski dense in $\mathbb{C}^n$, the above statements hold for almost all parameters $\mathbf{q} \in \mathbb{R}^m$, except that the number of real-valued isolated solutions varies as a function of $\mathbf{q}$ and is no longer a constant. For our applications, we only need the fact that the real roots of a polynomial system are finite and geometrically isolated for generic $\mathbf{q}$.

Our polynomial system (1.4) is parameterized by W which, in turns, is related to the given order-k tensor T, but is under-determined. In no way can the solutions be isolated. However, the solutions are structured into orbit varieties. By regarding (1.4) as a system over the quotient manifold $\mathscr{S}(p_1, q_1)/\mathscr{O}(q_1) \times \mathscr{S}(p_2, q_2)/\mathscr{O}(q_2)$, we have the same number of equations and unknowns (orbits). Lemma 3.6 can be applied, that is, there is a generic behavior of the solution set, interpreted as orbits, to (1.4) with respect to the parameter W.

COROLLARY 3.7. *For generic $W \in \mathbb{R}^{p_1 p_2 \times p_1 p_2}$, there are finitely many and geometrically isolated orbit varieties in the form (2.9) for the polynomial system (1.4).*

To complete our proof of convergence of the orbit $\{([Q_{1,[s]}], [Q_{2,[s]}])\}$, we recall the following result first proved in [3, Lemma 4.3].

LEMMA 3.8. *Let $\{a_s\}$ be a bounded sequence of real numbers with the property $|a_{s+1} - a_s| \to 0$ as $s \to \infty$. If the accumulation points for the sequence are isolated, then $\{a_s\}$ converges to a unique limit point.*

THEOREM 3.9. *For generic $W \in \mathbb{R}^{p_1 p_2 \times p_1 p_2}$, the sequence of orbits $\{([Q_{1,[s]}], [Q_{2,[s]}])\}$ whose representatives $\{(Q_{1,[s]}, Q_{2,[s])}\}$ are generated by (3.1) converge to a single limit point.*

Proof. The sequence $\{([Q_{1,[s]}], [Q_{2,[s]}])\}$ is clearly bounded, so it must have accumulation points. The accumulation points must satisfy the polynomial system (1.4). By Corollary (3.7), they are isolated for generic W. Replacing the role of $\{a_s\}$ in Lemma 3.8 by $\{[Q_{i,[s]}]\}$, $i = 1, 2$, together with the fact that $d_i([Q_{i,[s+1]}], [Q_{i,[s]}]) \to 0$ as $s \to \infty$, the sequence $\{[Q_{i,[s]}]\}$ converges to a unique limit point. □

The convergence of the sequence of orbits does not necessarily imply that the sequence of any representatives from the orbits converges. Our sequence $\{Q_{i,[s]}\}$, however, is a cross-section with the identity 1_{q_i} fixed across all orbits.

COROLLARY 3.10. *The sequence of iterates $\{(Q_{1,[s]}, Q_{2,[s]})\}$ generated by the scheme (3.1) converges.*

4. Conclusion. A mechanism needed for the simultaneous updating of two factor matrices in the low rank Tucker tensor approximation leads to a nonlinearly coupled matrix equation. The solution set is invariant under the orthogonal group action and, hence, consists of orbit variety. An iterative scheme analogous to the conventional power method for subspace iterations is proposed to track one cross-section of the orbits. The convergence analysis is much harder because the matrices used for iteration are themselves part of the dynamics. We resort to the notion of Riemannian manifolds and their quotients as a tool. In particular, we establish an isometric isomorphism (3.21) between the quotient space $\mathfrak{T}_Q\mathscr{S}(p, q)/\mathfrak{T}_Q[Q]$, which is the tangent space of the Stiefold manifold modulo the tangent space of one orbit, and the tangent space $\mathfrak{T}_{[Q]}(\mathscr{S}(p, q)/\mathscr{O}(q))$ of the Riemannian manifold $\mathscr{S}(p, q)/\mathscr{O}(q)$. Using the induced Riemannian metric, the convergence theory of the orbits as well as of the iterates is completed.

REFERENCES

[1] B. W. BADER AND T. G. KOLDA, *Efficient MATLAB computations with sparse and factored tensors*, SIAM J. Sci. Comput., 30 (2007/08), pp. 205–231.

[2] M. T. CHU AND N. T. TRENDAFILOV, *The orthogonally constrained regression revisited*, J. Comput. Graph. Statist., 10 (2001), pp. 746–771.

[3] M. T. CHU AND J. L. WATTERSON, *On a multivariate eigenvalue problem. I. Algebraic theory and a power method*, SIAM J. Sci. Comput., 14 (1993), pp. 1089–1106.

[4] H. DERKSEN AND G. KEMPER, *Computational invariant theory*, vol. 130 of Encyclopaedia of Mathematical Sciences, Springer, Heidelberg, enlarged ed., 2015. Invariant Theory and Algebraic Transformation Groups, VIII.

[5] M. HEIN AND T. BÜHLER, *An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA*, CoRR, abs/1012.0774 (2010).
[6] N. J. HIGHAM, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1160–1174.
[7] A. N. LANGVILLE AND W. J. STEWART, *The Kronecker product and stochastic automata networks*, J. Comput. Appl. Math., 167 (2004), pp. 429–447.
[8] A. RUHE, *Algorithms for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 10 (1973), pp. 674–689.
[9] A. J. SOMMESE AND C. W. WAMPLER, II, *The numerical solution of systems of polynomials arising in engineering and science*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.
[10] C. F. VAN LOAN, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123 (2000), pp. 85–100.
[11] ———, *Structured matrix problems from tensors*, in Exploiting Hidden Structure in Matrix Computations: Algorithms and Applications, vol. 2173 of Lecture Notes in Math., Springer, Cham, 2016, pp. 1–63.

5. Appendix. By now, our theory is complete for the general case. It might be illuminating to consider two special cases to further demonstrate the intriguing dynamics discussed above.

Example 2. The iterative scheme corresponding to the case $q_1 = q_2 = 1$ looks very much like the conventional power method

$$\left\{\begin{array}{rcl} \mathcal{A}(\mathbf{q}_{2,[s]}, \mathbf{q}_{2,[s]})\mathbf{q}_{1,[s]} & = & \mathbf{q}_{1,[s+1]}\lambda_{1,[s+1]}, \\ \mathcal{B}(\mathbf{q}_{1,[s+1]}, \mathbf{q}_{1,[s+1]})\mathbf{q}_{2,[s]} & = & \mathbf{q}_{2,[s+1]}\lambda_{2,[s+1]}, \end{array}\right.$$

where $\lambda_{1,[s]}, \lambda_{2,[s]}$ are simply the norms of the vector generated on the left hand side. The main difference is that the matrices $\mathcal{A}(\mathbf{q}_{2,[s]}, \mathbf{q}_{2,[s]})$ and $\mathcal{B}(\mathbf{q}_{1,[s+1]}, \mathbf{q}_{1,[s+1]})$ are not stationary until the iterates $(\mathbf{q}_{1,[s]}, \mathbf{q}_{2,[s]})$ reach the equilibrium point where the problem becomes an eigenvalue problem. The group action by $\mathscr{O}(1) = \{1, -1\}$ is simply a change of sign.

Example 3. Suppose that the operators $\mathcal{A} \in \mathbb{R}^{p_1 \times p_1}$ and $\mathcal{B} \in \mathbb{R}^{p_2 \times p_2}$ are invariant in s. The iterative scheme

$$\left\{\begin{array}{rcl} \mathcal{A}Q_{1,[s]} & = & Q_{1,[s+1]}P_{1,[s+1]} \\ \mathcal{B}Q_{2,[s]} & = & Q_{2,[s+1]}P_{2,[s+1]} \end{array}\right.$$

is equivalent to the simultaneous subspace iteration. Specifically, the bases of the new subspaces after $\mathcal{A}$ and $\mathcal{B}$ are applied to the subspaces spanned by the columns of $Q_{1,[s]}$ and $Q_{2,[s]}$, respectively, are represented by the orthonormal vectors of $Q_{1,[s+1]}$ and $Q_{2,[s+1]}$. Normally, the new bases would be obtained from the QR decomposition, but our scheme uses the bases from the polar decomposition. It is well known in the literature that the simultaneous q-dimensional subspace iteration will converge to the subspace spanned by the eigenvectors corresponding to the first q dominant eigenvalues. Even though the polar decomposition is used in place of the QR decomposition, the same convergence result prevails. The group action by orthogonal matrices simply changes one orthonormal basis to another one, but does not change the corresponding subspace.