

Low Rank Pure Quaternion Approximation for Pure Quaternion Matrices

Guangjing Song* Weiyang Ding† Michael K. Ng‡

January 1, 2021

Abstract

Quaternion matrices are employed successfully in many color image processing applications. In particular, a pure quaternion matrix can be used to represent red, green and blue channels of color images. A low-rank approximation for a pure quaternion matrix can be obtained by using the quaternion singular value decomposition. However, this approximation is not optimal in the sense that the resulting low-rank approximation matrix may not be pure quaternion, i.e., the low-rank matrix contains real component which is not useful for the representation of a color image. The main contribution of this paper is to find an optimal rank- r pure quaternion matrix approximation for a pure quaternion matrix (a color image). Our idea is to use a projection on a low-rank quaternion matrix manifold and a projection on a quaternion matrix with zero real component, and develop an alternating projections algorithm to find such optimal low-rank pure quaternion matrix approximation. The convergence of the projection algorithm can be established by showing that the low-rank quaternion matrix manifold and the zero real component quaternion matrix manifold has a non-trivial intersection point. Numerical examples on synthetic pure quaternion matrices and color images are presented to illustrate the projection algorithm can find optimal low-rank pure quaternion approximation for pure quaternion matrices or color images.

Keywords: Color images, pure quaternion matrices, low-rank approximation, manifolds

1 Introduction

The RGB color model is one of the most commonly applied additive color models. Each pixel using the RGB color model consists of three channels, i.e., red (R), green (G), and blue(B), which can be encoded on the three imaginary parts of a quaternion. The quaternion representation for color images has been proposed and widely employed in the literature [7, 9, 14, 20, 28, 29, 32]. It is a main advantage of the quaternion approach that color images can be studied and processed holistically as a vector field [9, 29, 32]. Tools and

*School of Mathematics and Information Sciences, Weifang University, Weifang 261061, P.R. China. Email: sgjshu@163.com

†Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, P.R. China. Research supported in part by NSFC Grant 11801479 and HKRGC GRF 12301619. Email: weiyang.ding@gmail.com

‡The Corresponding Author. Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong. Research supported in part by the HKRGC GRF 12200317, 12300218, 12300519 and 17201020. E-mail: mng@maths.hku.hk

methods for gray-scale image processing are extended to the color image processing field via the quaternion algebra theory and computation, such as the matrix completion [16], the Fourier transform [28, 29], the wavelet transform [10], the principal component analysis [41], and the dictionary learning algorithms [4, 38, 40].

In this paper, we also employ the quaternion matrix representation for an RGB color image and target at the optimal rank- r approximation that should also represent an RGB color image. Note that the representation quaternion matrix should be pure quaternion, i.e., admitting a zero real part, since we assign the red, green, and blue channels to the three imaginary parts respectively. Without the pure quaternion restriction, a low-rank approximation of any quaternion matrix can be obtained by the quaternion singular value decomposition (QSVD) [42], i.e., the quaternion counterpart for the Eckart-Young-Mirsky theorem. Nevertheless, this low-rank approximation may not be optimal in the sense that the approximation matrix is not necessarily a pure quaternion matrix. That is, the approximation matrix contains real component information which is useless for a color image. We take a 2-by-2 pure quaternion matrix for illustration. Denote

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \mathbf{j} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \mathbf{k}, \quad (1)$$

where \mathbf{i} , \mathbf{j} and \mathbf{k} are three imaginary units in the quaternion algebra. By applying the QSVD on \mathbf{A} , we can calculate an optimal rank-1 approximation as

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0 & -0.35 \\ 0 & 0.35 \end{pmatrix} + \begin{pmatrix} 0.85 & 0 \\ 0.35 & 0 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0 & -0.85 \\ 0 & -0.85 \end{pmatrix} \mathbf{j} + \begin{pmatrix} 0.35 & 0 \\ 0.85 & 0 \end{pmatrix} \mathbf{k}.$$

Obviously, the approximation matrix $\tilde{\mathbf{A}}$ contains real components. In this case, the pure quaternion part of $\tilde{\mathbf{A}}$

$$\tilde{\mathbf{A}}_p = \begin{pmatrix} 0.85 & 0 \\ 0.35 & 0 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0 & -0.85 \\ 0 & -0.85 \end{pmatrix} \mathbf{j} + \begin{pmatrix} 0.35 & 0 \\ 0.85 & 0 \end{pmatrix} \mathbf{k}$$

is usually chosen as the pure quaternion approximation of \mathbf{A} . However, $\text{rank}(\tilde{\mathbf{A}}_p) = 2$, which is saying that $\tilde{\mathbf{A}}_p$ is not an optimal rank-1 pure quaternion approximation of \mathbf{A} .

In color image processing, we are interested in finding an optimal fixed-rank pure quaternion approximation of a color image. Mathematically, it can be formulated as the following optimization problem

$$\begin{aligned} \min \quad & \|\mathbf{A} - \mathbf{X}\|_{\mathbb{F}}^2, \\ \text{s.t.} \quad & \text{rank}(\mathbf{X}) = r, \\ & \text{Re}(\mathbf{X}) = 0, \end{aligned} \quad (2)$$

where \mathbf{A} is a given pure quaternion matrix and $\text{Re}(\mathbf{X})$ denotes the real part of \mathbf{X} . In the literature, there are several algorithms [17, 18, 20, 26, 27] for computing the eigenvalues as well as the singular values of quaternion matrices. To our best knowledge, this is the first attempt to study the low-rank pure quaternion approximation problem.

1.1 The Contribution

Alternating projection method finds a point in the intersection of two closed convex sets by iteratively projecting a point first onto one set and then the other. It has a long history which can be traced back to John Von Neumann [35], where the alternating projections between two closed subspaces of a Hilbert space is guaranteed to globally converge to a intersection

point of the two subspaces, if they intersect non-trivially. Alternating projection method has been used in a wide range of some classical mathematics problems as well as engineering applications because of it is a gradient-free method (without requiring tuning the step size) and usually has fast speed of convergence. See, e.g., convex feasibility problem [5], signal processing [8], finance [15], machine learning [37], and so on ([2, 12, 13, 19, 22, 23, 33, 39] and references therein).

By contrast, alternating projection method on nonconvex sets are remained rather undeveloped. Cadzow [6] showed the convergence of an alternating projections scheme by the Zangwill's global convergence theorem. Lewis and Malich [24] further showed that alternating projection method converges locally at a linear rate when two manifolds intersect transversally. Later, Fredrik and Marcus [3] generalized the "transversally" intersecting condition as the "nontangentially" intersecting condition.

In this paper, we consider the alternating projection method on two manifolds: one is a fixed-rank quaternion matrix manifold and the other is a zero real component quaternion matrix manifold. The convergence results for the alternating projections algorithm on these two manifolds are derived. Furthermore, we propose an initialization strategy to make the alternating projection method more practical and reliable. Numerical examples on synthetic pure quaternion matrices and color images are presented to illustrate the performances of the proposed algorithm.

The rest of this paper is organized as follows. In Section 2, we summarize some notations used throughout this paper. The preliminaries for quaternion matrix and manifolds theory are presented. We also study the intersection of the low-rank quaternion matrix manifold and the zero real component quaternion matrix manifold. In Section 3, an alternating projections algorithm is given and proved to linearly converge to a good approximation. We further propose an initialization strategy in Section 4 to make the alternating projection method more practical and reliable. In Section 5, we conduct some numerical experiments to demonstrate our theoretical results. Some concluding remarks are given in Section 6.

2 Mathematical Preliminaries

2.1 Quaternion and Quaternion Matrix

A quaternion number consists of one real part and three imaginary parts given by

$$\mathbf{q} = q_r + q_i \mathbf{i} + q_j \mathbf{j} + q_k \mathbf{k}$$

where $q_r, q_i, q_j, q_k \in \mathbb{R}$ and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are three imaginary units. Throughout this paper, any boldface symbol indicates a quaternion number, vector, or matrix. We use \mathbb{H} to denote the quaternion algebra. The quaternion \mathbf{q} is called a pure quaternion if it has a zero real part, i.e., $q_r = 0$. The conjugate and modulus of \mathbf{q} are respectively defined by

$$\mathbf{q}^* = q_r - q_i \mathbf{i} - q_j \mathbf{j} - q_k \mathbf{k} \quad \text{and} \quad |\mathbf{q}| = \sqrt{q_r^2 + q_i^2 + q_j^2 + q_k^2}.$$

Quaternions form a division algebra when equipped with the componentwise addition, the componentwise scalar multiplication over \mathbb{R} , and the Hamilton product given by

- $\mathbf{i} \cdot 1 = 1 \cdot \mathbf{i} = \mathbf{i}, \mathbf{j} \cdot 1 = 1 \cdot \mathbf{j} = \mathbf{j}, \mathbf{k} \cdot 1 = 1 \cdot \mathbf{k} = \mathbf{k};$
- $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \mathbf{ik} = -\mathbf{ki} = \mathbf{j}.$

Note that the quaternion multiplication is noncommutative, i.e., $\mathbf{p}\mathbf{q}$ may not equal $\mathbf{q}\mathbf{p}$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{H}$.

A quaternion matrix is represented by $\mathbf{A} = A_0 + A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k} \in \mathbb{H}^{m \times n}$ and the real part of \mathbf{A} is denoted by $\text{Re}(\mathbf{A}) = A_0$. The identity quaternion matrix \mathbf{I} is the same as the classical identity matrix. The inverse \mathbf{B} of a quaternion matrix \mathbf{A} exists if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. A quaternion matrix is unitary if $\mathbf{A}^*\mathbf{A} = \mathbf{AA}^* = \mathbf{I}$, where \mathbf{A}^* is the conjugate transpose of \mathbf{A} . An alternative approach to handle an $m \times n$ quaternion matrix is to consider the subset of the ring $\mathbb{R}^{4m \times 4n}$:

$$\Gamma := \left\{ X = \begin{pmatrix} A_0 & -A_1 & A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ -A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n} : A_0, A_1, A_2, A_3 \in \mathbb{R}^{m \times n} \right\}.$$

Inheriting the operations on $\mathbb{R}^{4m \times 4n}$, Γ actually forms a subring. Define a mapping ϕ as

$$\phi : \mathbf{A} = A_0 + A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k} \in \mathbb{H}^{m \times n} \mapsto \hat{\mathbf{A}} = \begin{pmatrix} A_0 & -A_1 & A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ -A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix} \in \Gamma. \quad (3)$$

Then ϕ is a bijection and preserves the operations, which guarantees that the quaternion matrix set $\mathbb{H}^{m \times n}$ and the real matrix set Γ are essentially the same (Remark 2.1 in [42]). In addition, $\hat{\mathbf{A}}$ in (3) is called the real representation of \mathbf{A} .

Due to the non-commutative nature of quaternion multiplication, the linear independence of a set of vectors over \mathbb{H}^n can be defined as right and left linear independence, respectively. $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{H}^n$ are said to be right (left) linear independent if there dose not exist nonzero quaternions $\mathbf{a}_1, \dots, \mathbf{a}_n$ such that

$$\mathbf{v}_1\mathbf{a}_1 + \dots + \mathbf{v}_n\mathbf{a}_n = 0 \quad (\mathbf{a}_1\mathbf{v}_1 + \dots + \mathbf{a}_n\mathbf{v}_n = 0).$$

Based on the definition of right independence of quaternion vectors, we can introduce the rank of a quaternion matrix.

Definition 2.1 ([42]). *The maximum number of right linearly independent columns of a quaternion matrix $\mathbf{A} \in \mathbb{H}^{n_1 \times n_2}$ is called the rank of \mathbf{A} .*

Nevertheless, the rank could be different from the maximum number of left linearly independent columns or right linearly independent rows since the quaternion multiplication is noncommutative. Moreover, the singular value decomposition of a quaternion matrix reveals the rank of the quaternion matrix which can be given as follows.

Theorem 2.1 (QSVD [42]). *Let $\mathbf{A} \in \mathbb{H}^{n_1 \times n_2}$ with $\text{rank}(\mathbf{A}) = r$. Then there exist two unitary quaternion matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_1}] \in \mathbb{H}^{n_1 \times n_1}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_2}] \in \mathbb{H}^{n_2 \times n_2}$ such that*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad (4)$$

where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in \mathbb{R}^{n_1 \times n_2}$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ are positive singular values of \mathbf{A} .

Compared with the real and complex cases, the singular values of a quaternion matrix are still real numbers, while the two unitary matrices are quaternion matrices. Similar to the complex matrix case, an optimal rank- r approximation of \mathbf{A} can be given as $\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^*$.

2.2 Manifolds

A smooth manifold is a pair (\mathcal{M}, φ) , where \mathcal{M} is a topological manifold and φ is a smooth structure on \mathcal{M} . Let $\gamma : \mathbb{R} \rightarrow \mathcal{M}, t \rightarrow \gamma(t)$ be a smooth curve in \mathcal{M} with $\gamma(0) = x$, and $\Upsilon(x)$ be the set of all real-valued functions $f : \mathcal{M} \rightarrow \mathbb{R}$ which are smooth around $x \in \mathcal{M}$. Then the mapping

$$v_x : \Upsilon(x) \rightarrow \mathbb{R}, f \rightarrow v_x f := \left. \frac{df(\gamma(t))}{dt} \right|_{t=0}$$

is called the tangent vector of x to the curve γ at $t = 0$. The set of all derivations of \mathcal{M} at x is a vector space called the tangent space to \mathcal{M} at x , and is denoted by $\mathcal{T}_{\mathcal{M}}(x)$.

Some well known matrix manifolds in the literature are the orthogonal Stiefel manifold, the Grassmann manifold and the fixed rank matrix manifold which can be constructed from $\mathbb{R}^{m \times n}$ by taking embedded or quotient operations. In order to better understand manifolds with its related definitions (e.g., charts, atlases and tangent spaces) and some optimization algorithms on matrix manifolds, we refer to [1, 21] and the references therein. Here, we focus on the fixed rank matrix manifold whose dimension and tangent space are given as follows.

Lemma 2.2 (Proposition 2.1 in [34]). *Let \mathcal{M} be the set of $m \times n$ real matrices with a fixed rank r . Then \mathcal{M} is an embedded manifold of $\mathbb{R}^{m \times n}$ with dimension $(m + n - r)r$. Suppose that the skinny SVD of $X \in \mathcal{M}$ is given by $X = U\Sigma V^T$, with $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ being two column unitary matrices. Its tangent space $\mathcal{T}_{\mathcal{M}}(X)$ at X is given by*

$$\mathcal{T}_{\mathcal{M}}(X) = \left\{ [U, U_{\perp}] \begin{pmatrix} \mathbb{R}^{r \times r} & \mathbb{R}^{r \times (n-r)} \\ \mathbb{R}^{(m-r) \times r} & \mathbb{0}^{(m-r) \times (n-r)} \end{pmatrix} [V, V_{\perp}]^T \right\},$$

where U_{\perp} and V_{\perp} are the unitary complements of U and V , respectively.

In this section, our aim is to show the intersection of the fixed rank quaternion matrix set and the zero real component quaternion matrix set is a manifold, which provides a theoretical guarantee of the local linear convergence for the alternating projection method proposed in the next section. The angle between two manifolds plays an important role in checking the alternating projection method can be applied or not, then we need to introduce the following definitions first.

Definition 2.2 (Definition 3.1 in [3]). *Suppose that \mathcal{M}_1 and \mathcal{M}_2 are two manifolds, then given $A \in \mathcal{M}_1 \cap \mathcal{M}_2$, set*

$$F_j^{\xi} = \{B_j \in \mathcal{M}_j \setminus A, \|B_j - A\| \leq \xi \text{ and } B_j - A \perp T_{\mathcal{M}_1 \cap \mathcal{M}_2}(A)\}, j = 1, 2.$$

If $F_j^{\xi} \neq \emptyset$, $j = 1, 2$, for all $\xi > 0$, we define the angle $\alpha(A)$ of \mathcal{M}_1 and \mathcal{M}_2 at A as

$$\alpha(A) = \cos^{-1}(\sigma(A)),$$

where

$$\sigma(A) = \lim_{\xi \rightarrow 0} \sup_{B_1 \in F_1^{\xi}, B_2 \in F_2^{\xi}} \left\{ \frac{\langle B_1 - A, B_2 - A \rangle}{\|B_1 - A\| \|B_2 - A\|} \right\}.$$

Definition 2.3 (Definition 3.3 in [3]). *Suppose that \mathcal{M}_1 and \mathcal{M}_2 are two manifolds. Points $A \in \mathcal{M}_1 \cap \mathcal{M}_2$ where the angle is defined will be called nontrivial intersection points. For such points, we say that A is tangential if $\alpha(A) = 0$ and non-tangential if $\alpha(A) > 0$.*

For quaternion matrices, we can also show the fixed-rank ones form a manifold. The proof can be found in Appendix.

Lemma 2.3. *Denote $\mathcal{Q} := \{\mathbf{E} = E_0 + E_1\mathbf{i} + E_2\mathbf{j} + E_3\mathbf{k} \in \mathbb{H}^{m \times n}, \text{rank}(\mathbf{E}) = r\}$. Then \mathcal{Q} is an embedded submanifold of $\mathbb{R}^{4m \times 4n}$ with dimension $4(m + n - r)r$.*

By the real representation of a quaternion matrix given in (3), we can get \mathcal{Q} is isomorphic to the set of structured real matrices:

$$\mathcal{M}_r := \left\{ X = \begin{pmatrix} A_0 & -A_1 & A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ -A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}, \text{rank}(X) = 4r \right\}, \quad (5)$$

and the set of pure quaternion matrices is isomorphic to the set of structured real matrices:

$$\mathcal{M}_* := \left\{ X = \begin{pmatrix} 0 & -A_1 & A_2 & -A_3 \\ A_1 & 0 & -A_3 & A_2 \\ -A_2 & A_3 & 0 & -A_1 \\ A_3 & -A_2 & A_1 & 0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n} \right\}, \quad (6)$$

respectively. Obviously, \mathcal{M}_* is a linear subspace, thus also a manifold. After that the main task of this section can be rewritten as proving

$$\mathcal{M}_{r*} := \mathcal{M}_r \cap \mathcal{M}_* = \left\{ X = \begin{pmatrix} 0 & -A_1 & A_2 & -A_3 \\ A_1 & 0 & -A_3 & A_2 \\ -A_2 & A_3 & 0 & -A_1 \\ A_3 & -A_2 & A_1 & 0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}, \text{rank}(X) = 4r \right\} \quad (7)$$

is a manifold. Moreover, \mathcal{M}_r and \mathcal{M}_* can be seen as the vanishing of different polynomial sets over \mathbb{R}^n , which motivated us to apply algebraic geometry methods to study the above problem. Before moving on, we need to introduce the following definitions and results which are needed in the sequel.

Theorem 2.4 (Theorem 2.1 in [36]). *Given a real algebraic variety $\mathcal{V} \in \mathbb{R}^n$, we can write $\mathcal{V} = \bigcup_{j=0}^m \mathcal{V}_j$ where each \mathcal{V}_j is either void or a $\mathbb{C}^{(\infty)}$ -manifold of dimension j . Moreover, each \mathcal{V}_j contains at most a finite number of connected components.*

Theorem 2.4 shows us that the main part of a variety is a manifold. For a given real algebraic variety $\mathcal{V} \in \mathbb{R}^n$, if we identify \mathbb{R}^n as a subset of \mathbb{C}^n and denote $\mathbb{I}_{\mathbb{R}}(\mathcal{V})$ as the set of real polynomials that vanish on \mathcal{V} , then \mathcal{V} has a related complex variety given by its Zariski closure

$$\mathcal{V}_{Zar} = \{z \in \mathbb{C}^n : p(z) = 0, \forall p \in \mathbb{I}_{\mathbb{R}}(\mathcal{V})\},$$

which is defined as the subset in \mathbb{C}^n of common zeros to all polynomials that vanish on \mathcal{V} . A given real algebraic variety \mathcal{V} is called irreducible if there does not exist any non-trivial decompositions of the form $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$, where \mathcal{V}_1 and \mathcal{V}_2 are real algebraic varieties. A point $z \in \mathcal{V}$ is non-singular if it is non-singular in the sense of algebraic geometry as an element of \mathcal{V}_{Zar} . Denote ∇ as the gradient operator and set $\mathcal{N}_{\mathcal{V}}(z) = \{\nabla p(z) : p \in \mathbb{I}_{\mathbb{R}}(\mathcal{V})\}$. The set of non-singular points in \mathcal{V} is denoted as \mathcal{V}^{ns} . The following results provide some criteria to check whether a point is a non-singular point or not.

Lemma 2.5 (Proposition 6.3 in [3]). *Let $\mathcal{V} \in \mathbb{R}^n$ be a irreducible real algebraic variety of dimension m . Then $\dim \mathcal{N}_{\mathcal{V}}(z) \leq n - m$ for all $z \in \mathcal{V}$ and z is non-singular if and only if $\dim \mathcal{N}_{\mathcal{V}}(z) = n - m$.*

Lemma 2.6 (Proposition 6.4 in [3]). *Let \mathcal{V} be an irreducible real algebraic variety of dimension m . Then the decomposition $\mathcal{V} = \bigcup_{j=1}^m \mathcal{M}_j$ in Theorem 2.4 can be chosen as $\mathcal{V}^{ns} = \mathcal{M}_m$.*

Lemma 2.7 (Proposition 6.5 in [3]). *Suppose that \mathcal{V}_1 and \mathcal{V}_2 are irreducible real algebraic varieties and that $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2$ is irreducible and strictly smaller than both \mathcal{V}_1 and \mathcal{V}_2 . Then each point in $\mathcal{V}_1^{ns} \cap \mathcal{V}_2^{ns} \cap \mathcal{V}^{ns}$ is a non-trivial intersection point.*

Lemma 2.8 (Theorem 6.6 in [3]). *Suppose that \mathcal{V}_1 and \mathcal{V}_2 are irreducible real algebraic varieties and that $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2$ is irreducible and strictly smaller than both \mathcal{V}_1 and \mathcal{V}_2 . Let the dimension of \mathcal{V} is m . If $\mathcal{V}^{ns, nt} \neq \emptyset$, then $\mathcal{V} \setminus \mathcal{V}^{ns, nt}$ is a real algebraic variety of dimension strictly less than m . A sufficient condition for this to happen is that there exist a point $A \in \mathcal{V}_1^{ns} \cap \mathcal{V}_2^{ns}$ such that*

$$\dim(T_{\mathcal{V}_1^{ns}}(A) + T_{\mathcal{V}_2^{ns}}(A)) \leq m.$$

In practice, we need to check a given variety is irreducible or not, thus the following results are needed.

Definition 2.4 (Definition 6.7 in [3]). *Suppose we are given a number $j \in \mathbb{N}$ and an index set I such that for each $i \in I$, there exist an open connected $\Omega_i \subseteq \mathbb{R}^j$ and a real analytic map $\phi_i : \Omega_i \rightarrow \mathcal{V}$. Then \mathcal{V} is said to be covered with analytic patches, if for each $A \in \mathcal{V}$, there exists an $i \in I$ and a radius r_A such that*

$$\mathcal{V}_r \cap \text{Ball}_{\mathbb{R}^n}(A, r_A) = \text{Im} \phi_i \cap \text{Ball}_{\mathbb{R}^n}(A, r_A).$$

Lemma 2.9 (Proposition 6.8 in [3]). *Let \mathcal{V} be a real algebraic variety. If \mathcal{V} is connected and can be covered with analytic patches, then \mathcal{V} is irreducible.*

The following lemma show us a method to compute the dimension of a variety.

Lemma 2.10 (Proposition 6.9 in [3]). *Under the assumption of Lemma 2.9, suppose in addition that an open subset of \mathcal{V} is the image of a bijective real analytic map defined on a subset of \mathbb{R}^d . Then \mathcal{V} has dimension d .*

For an arbitrary quaternion matrix $\mathbf{A} = A_0 + A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k} \in \mathbb{H}^{m \times n}$, denote

$$\mathcal{V}_{sr} := \left\{ X = \begin{pmatrix} A_0 & -A_1 & A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ -A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}, \text{rank}(X) \leq 4r \right\}. \quad (8)$$

Then we can derive the following results.

Theorem 2.11. *Let \mathcal{M}_* and \mathcal{V}_{sr} be given as in (6) and (8), respectively. Then \mathcal{M}_* is an affine subspaces of dimension $3mn$, \mathcal{V}_{sr} is an irreducible real algebraic variety of dimension $4(m+n)r - 4r^2$, and*

$$\mathcal{V} := \mathcal{V}_{sr} \cap \mathcal{M}_* = \left\{ X = \begin{pmatrix} 0 & -A_1 & A_2 & -A_3 \\ A_1 & 0 & -A_3 & A_2 \\ -A_2 & A_3 & 0 & -A_1 \\ A_3 & -A_2 & A_1 & 0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}, \text{rank}(X) \leq 4r \right\} \quad (9)$$

is an irreducible algebraic variety of dimension $3(m+n)r - 3r^2$.

Proof. First, denote

$$\Gamma := \left\{ X = \begin{pmatrix} A_0 & -A_1 & A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ -A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}, A_i, i = 0, 1, 2, 3 \in \mathbb{R}^{m \times n} \right\}.$$

Obviously, Γ is a linear space of dimension $4mn$. The set \mathcal{M}_* is obtained by adding the constraint $A_0 = 0$ to Γ . Thus, it is an affine space with dimension $3mn$.

Second, we will show that \mathcal{V}_{sr} is an irreducible real algebraic variety with dimension $4(m+n)r - 4r^2$. It is well known that a matrix has rank r if and only if there exists at least a non-zero $r \times r$ nonsingular minor which is a matrix obtained by deleting $n-r$ rows and columns and all the $(r+1) \times (r+1)$ minors are zeros. Then if a matrix in Γ has rank $4r$ then there exists at least a non-zero $4r \times 4r$ invertible minor and all $(4r+1) \times (4r+1)$ minors are zero. Hence, \mathcal{V}_{sr} is the variety induced by the determinants of these minors. By the quaternion singular value decomposition given in [42], any $A \in \Gamma$ with $\text{rank}(A) \leq 4r$ can be factorized into

$$A = USV^T = \begin{pmatrix} U_0 & -U_1 & U_2 & -U_3 \\ U_1 & U_0 & -U_3 & U_2 \\ -U_2 & U_3 & U_0 & -U_1 \\ U_3 & -U_2 & U_1 & U_0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 & 0 & 0 \\ 0 & \Sigma & 0 & 0 \\ 0 & 0 & \Sigma & 0 \\ 0 & 0 & 0 & \Sigma \end{pmatrix} \begin{pmatrix} V_0 & -V_1 & V_2 & -V_3 \\ V_1 & V_0 & -V_3 & V_2 \\ -V_2 & V_3 & V_0 & -V_1 \\ V_3 & -V_2 & V_1 & V_0 \end{pmatrix}^T, \quad (10)$$

with $U_i \in \mathbb{R}^{m \times r}$, $V_i \in \mathbb{R}^{n \times r}$ ($i = 0, 1, 2, 3$) and $\Sigma \in \mathbb{R}^{r \times r}$ being a diagonal matrix. In the other hand, if a matrix A can be expressed as (10), then $\text{rank}(A) \leq 4r$, i.e., $A \in \mathcal{V}_{sr}$. We see that \mathcal{V}_{sr} is connected and can be covered with one real polynomial. Then \mathcal{V}_{sr} is irreducible by Lemma 2.9.

Next, choose a subset of \mathcal{V}_{sr} with the diagonal elements of Σ being nonzero and different with each other. Then the freedom of the column unitary matrices $U \in \mathbb{R}^{4m \times 4r}$ and $V \in \mathbb{R}^{4n \times 4r}$ in (10) are $4mr - \frac{r(4r+1)}{2}$ and $4nr - \frac{r(4r+1)}{2}$, respectively. And the freedom of the diagonal matrix $S \in \mathbb{R}^{4r \times 4r}$ is r . Thus, the subsets of such matrices can be identified with $\mathbb{R}^{4mr - \frac{r(4r+1)}{2}}$, $\mathbb{R}^{4nr - \frac{r(4r+1)}{2}}$ and \mathbb{R}^r , respectively. Denote the inverses of the identification by

$$\iota_1 : \mathbb{R}^{4mr - \frac{r(4r+1)}{2}} \rightarrow \mathbb{R}^{4m \times 4r}; \quad \iota_2 : \mathbb{R}^r \rightarrow \mathbb{R}^{4r \times 4r}; \quad \iota_3 : \mathbb{R}^{4nr - \frac{r(4r+1)}{2}} \rightarrow \mathbb{R}^{4n \times 4r}; \quad (11)$$

and denote $\Omega \subset \mathbb{R}^{4(m+n)r - 4r^2}$ as the open set corresponding to those matrices with Σ possessing different diagonal elements. Define $\phi : \Omega \rightarrow \mathcal{V}_{sr}$ by

$$\phi(y_1, y_2, y_3) = \iota_1(y_1) \cdot \iota_2(y_2) \cdot (\iota_3(y_3))^T. \quad (12)$$

It is easy to see that ϕ is a polynomial and moreover a bijective correspondence with an open set Ω . Thus it follows Lemma 2.10 that the dimension of \mathcal{V}_{sr} is $4(m+n)r - 4r^2$.

Third, we turn our attention to $\mathcal{V} = \mathcal{V}_{sr} \cap \mathcal{M}_*$. Note that \mathcal{V} is obtained by adding the algebraic equations

$$U_0 \Sigma V_0^T - U_1 \Sigma V_1^T - U_2 \Sigma V_2^T - U_3 \Sigma V_3^T = 0 \quad (13)$$

to those entries of matrices in defining \mathcal{V}_{sr} , then it is also a real algebraic variety.

Then, we will apply Lemma 2.9-2.10 to show \mathcal{V} is an irreducible real algebraic variety with dimension $3(m+n)r - 3r^2$.

Let $U \in \mathbb{R}^{4m \times 4r}$, $V \in \mathbb{R}^{4n \times 4r}$ and $S \in \mathbb{R}^{4r \times 4r}$ be defined as (10). We set all the elements of U_0 as undetermined variables and other values are fixed. Then the $m \times n$ linear equations

in (13) relate to the undetermined variables $(U_0)_{i,j}, i = 1, \dots, m, j = 1, \dots, r$ may have 0, 1 or infinite solutions (the number of solutions was decided by the property of V, Σ , and the other variables of U). Suppose the remaining values are chosen such that the equations have a unique solution relate to every undermined variable $(U_0)_{i,j}, i = 1, \dots, m, j = 1, \dots, r$, respectively. Denote the corresponding matrix by \hat{U} , after these $(U_0)_{i,j}, i = 1, \dots, m, j = 1, \dots, r$ are fixed. Then, a real analytic mapping θ from \hat{U}, S and V to \mathcal{V} can be constructed as follows:

$$\begin{aligned} & \theta((U_1)_{1,1}, \dots, (U_1)_{m,r}, \dots, (U_3)_{1,1}, \dots, (U_3)_{m,r}, \sigma_1, \dots, \sigma_r, (V_0)_{1,1}, \dots, (V_3)_{n,r}) \\ &= \begin{pmatrix} U_0 & -U_1 & U_2 & -U_3 \\ U_1 & U_0 & -U_3 & U_2 \\ -U_2 & U_3 & U_0 & -U_1 \\ U_3 & -U_2 & U_1 & U_0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 & 0 & 0 \\ 0 & \Sigma & 0 & 0 \\ 0 & 0 & \Sigma & 0 \\ 0 & 0 & 0 & \Sigma \end{pmatrix} \begin{pmatrix} V_0 & -V_1 & V_2 & -V_3 \\ V_1 & V_0 & -V_3 & V_2 \\ -V_2 & V_3 & V_0 & -V_1 \\ V_3 & -V_2 & V_1 & V_0 \end{pmatrix}^T \\ &= \begin{pmatrix} W_0 & -W_1 & W_2 & -W_3 \\ W_1 & W_0 & -W_3 & W_2 \\ -W_2 & W_3 & W_0 & -W_1 \\ W_3 & -W_2 & W_1 & W_0 \end{pmatrix}, \end{aligned} \quad (14)$$

with

$$\begin{aligned} W_0 &= U_0 \Sigma V_0^T - U_1 \Sigma V_1^T - U_2 \Sigma V_2^T - U_3 \Sigma V_3^T, W_1 = -U_0 \Sigma V_1^T - U_1 \Sigma V_0^T + U_2 \Sigma V_3^T - U_3 \Sigma V_2^T, \\ W_2 &= U_0 \Sigma V_2^T + U_1 \Sigma V_3^T - U_2 \Sigma V_0^T - U_3 \Sigma V_1^T, W_3 = -U_0 \Sigma V_3^T - U_1 \Sigma V_2^T - U_2 \Sigma V_1^T - U_3 \Sigma V_0^T. \end{aligned}$$

Note that the entries of W_0 in (14) can be zeros when the variables in \hat{U}, S and V are chosen as above which can guarantee equations in (13) are satisfied. It is saying that \mathcal{V} is the image of θ . Let Γ be a particular connected component of (\hat{U}, Σ, V) . We establish a function ψ with Γ as follows:

$$\psi_\Gamma(y) = U(y) \Sigma(y) V(y)^T, \quad y \in \Gamma. \quad (15)$$

Denote \mathbb{I} as the set of all possible π and Γ . It can be found that for each matrix in \mathcal{V} is in the image of at least one ψ_Γ where $\Gamma \in \mathbb{I}$. Then by Definition 2.4, \mathcal{V} can be covered by $\{\psi_\Gamma\}_{\Gamma \in \mathbb{I}}$.

Furthermore, in order to show \mathcal{V} is irreducible we need to show \mathcal{V} is connected. It is sufficient to prove \mathcal{V} is path connected, i.e., for any two matrices $A, B \in \mathcal{V}$, there exist a continuous map f from the unit interval $[0, 1]$ to \mathcal{V} such that $f(0) = A$ and $f(1) = B$. Without loss of generality, we show that for an arbitrary $A \in \mathcal{V}_{rn}$, it is connected with

$$X = \begin{pmatrix} \hat{\mathbf{1}} & \hat{\mathbf{1}} & \cdots & \hat{\mathbf{1}} \\ \hat{\mathbf{1}} & \hat{\mathbf{1}} & \cdots & \hat{\mathbf{1}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{1}} & \hat{\mathbf{1}} & \cdots & \hat{\mathbf{1}} \end{pmatrix} \quad \text{with } \hat{\mathbf{1}} = \begin{pmatrix} 0 & -1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \end{pmatrix}$$

instead. Suppose that $A, B \in \mathcal{V}$ are arbitrary and path connected with the X matrix, respectively. Thus, there are two continuous maps f and g which are from the unit interval $[0, 1]$ to \mathcal{V} with $f(0) = A, f(1) = X, g(0) = X$ and $g(1) = B$. Setting $\tau(x) = (1 - x)f(x) + xg(x)$, it is easy to see that $\tau(x)$ is continuous and satisfying $\tau(0) = f(0) = A$ and $\tau(1) = g(1) = B$. Then \mathcal{V} is path connected. Let A be fixed. We assume that the diagonal elements of Σ are nonnegative and decreasingly ordered and $\Sigma_{ii} = 1, i = 1, 2, 3, 4$. Pick σ such that $\sigma(i) = \sigma(j) = k$ for all i, j and choose Ω such that the representation is in the

form (12). If the second diagonal value in Σ is negative, then we continuously change it to the positivity inside Ω . Then the values of y corresponding to columns 1 through 8 of U can be continuously moved until all elements of the columns 1 through 4 as $[\mathbf{1}, \mathbf{1}, \dots, \mathbf{1}]^T$ with

$$\mathbf{1} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \end{pmatrix}.$$

At this point, all values of U except the first fourth columns vanish, increasing the first value of each row whenever necessary to stay inside Ω . We can move y so that the columns 1 though 4 of V can be written as $[\mathbf{a}, \mathbf{a}, \dots, \mathbf{a}]^T$ which satisfies $\mathbf{1} \cdot \mathbf{a} = \hat{\mathbf{1}}$. Thus, the matrix $\hat{\mathbf{1}}$ can be arrived which is saying that \mathcal{V} is connected.

In the end, we need to determine the dimension of \mathcal{V} . Consider again the map introduced earlier as (12), with the difference that the diagonal blocks are zeros. In order to guarantee (13) is satisfied, $(m+n)r - r^2$ additional constraints are added on these variables. It is naturally to define a real analytic map on the open subset Ξ of $\mathbb{R}^{3(m+n)r-3r^2}$. By (10), there exist three matrices U, V and S such that $A = USV^T$. The sets of $U \in \mathbb{R}^{4m \times 4r}$, $S = \text{diag}\{\Sigma, \Sigma, \Sigma, \Sigma\} \in \mathbb{R}^{4r \times 4r}$ and $V \in \mathbb{R}^{4n \times 4r}$ contain $3mr - \frac{r(3r+1)}{2}$, r , and $3nr - \frac{r(3r+1)}{2}$ independent variables, respectively. Therefore, U, S and V identify the set of matrices with $\mathbb{R}^{3mr - \frac{r(3r+1)}{2}}$, \mathbb{R}^r and $\mathbb{R}^{3nr - \frac{r(3r+1)}{2}}$, respectively. Then we can identify the set of such matrices with $\mathbb{R}^{3(m+n)r-3r^2}$. Denote the inverse of the identification as

$$\iota_1 : \mathbb{R}^{3mr - \frac{r(3r+1)}{2}} \rightarrow \mathbb{R}^{4m \times 4r}; \quad \iota_2 : \mathbb{R}^r \rightarrow \mathbb{R}^{4r \times 4r}; \quad \iota_3 : \mathbb{R}^{3nr - \frac{r(3r+1)}{2}} \rightarrow \mathbb{R}^{4n \times 4r};$$

and $\Omega \subset \mathbb{R}^{3(m+n)r-3r^2}$ as the open set corresponding to matrices with this structure. Ψ is a bijection with an open subset of \mathcal{V} . Hence, by Lemma 2.10 we can derive the dimension of \mathcal{V} is $3(m+n)r - 3r^2$. \square

Moreover, we can get the following results.

Theorem 2.12. *Suppose that \mathcal{M}_r , \mathcal{M}_* and \mathcal{V}_{sr} are defined as (5), (6) and (8), then $\mathcal{V}_{sr}^{ns} = \mathcal{M}_r$ and*

$$\dim(\text{span}(T_{\mathcal{M}_r}(A) \cap T_{\mathcal{M}_*}(A)) \leq 3(m+n)r - 3r^2. \quad (16)$$

Proof. Recall Lemma 2.5 and Theorem 2.11, we only need to show

$$\dim N_{\mathcal{V}_{sr}}(A) = 4mn - (4(m+n)r - 4r^2) = 4(mn - mr - nr + r^2)$$

if and only if $\text{rank}(A) = 4r$. It follows Lemma 2.5 that $\dim N_{\mathcal{V}_{sr}}(A) \leq 4(mn - mr - nr + r^2)$, then it is sufficient to show this inequality is strict if $\text{rank}(A) < 4r$ and the reverse inequality holds when $\text{rank}(A) = 4r$. In this proof, the particular identification of Γ given in Theorem 2.11 with \mathbb{R}^{4mn} is important. Given a polynomial $p \in \mathbb{I}_{\mathcal{V}_{rs}}$ (where $\mathbb{I}_{\mathcal{V}_{rs}}$ is defined as the set of real polynomials that vanish on \mathcal{V}_{rs}) and two unitary matrices U and V with proper orders such that $q_{U,V}(X) = p(UXV^T)$ is clearly also in $\mathbb{I}_{\mathcal{V}_{rs}}$. Due to the particular choice of ω , we have $\nabla_{q_{(U,V)}}(B) = U\nabla_p(UXV^T)V^T$. Let A be fixed of $\text{rank}(A) = j \leq 4r$. Then there exist two unitary matrices \hat{U} and \hat{V} such that $\hat{U}A\hat{V}^T = S_j = \text{diag}\{\Sigma, \Sigma, \Sigma, \Sigma\}$, where Σ is a diagonal matrix whose diagonal are $\sigma_i, i = 1, \dots, j$ and 0 elsewhere. It follows that $\nabla_{q_{(\hat{U}, \hat{V})}}(A) = \hat{U}\nabla p(S_j)\hat{V}^T$, which implies that $\dim N_{\mathcal{V}_{sr}}(A) = \dim N_{\mathcal{V}_{sr}}(S)$. Then all $\mathbb{R}_{4r+4, 4r+4}$ (in order to keep the structure of \mathcal{V}_{sr}) subdeterminants of Γ form polynomials in $\mathbb{I}_{\mathcal{V}_{sr}}$. We can get $\dim N_{\mathcal{V}_{sr}}(S) \geq 4(mn - mr - nr + r^2)$, then it prove that any matrix X

in the set Γ with $\text{rank}(X) = 4r$ element of \mathcal{V}_{sr} is non-singular. In other direction, if $j < r$, similarly as the above we can construct a variety with dimension of $4(mn - mj - nj + j^2)$ which is bigger than $4(mn - mr - nr + r^2)$. Consider two fixed matrices $\tilde{U} \in \mathbb{R}^{4m \times 4}$ and $\tilde{V} \in \mathbb{R}^{4m \times 4}$ and define the map $\theta_{\tilde{U}, \tilde{V}} : \mathbb{R}^{4m \times 4} \rightarrow \tilde{\mathcal{V}}_{sr}$ via $\theta_{\tilde{U}, \tilde{V}}(x) = S + x\tilde{U}\tilde{V}^T$. Then

$$\text{span} \left\{ \frac{d}{dx} \theta_{\tilde{U}, \tilde{V}}(0) : \tilde{U}, \tilde{V} \in \mathbb{R}^{4m \times 4} \right\} = \Gamma,$$

which is saying that the dimension of the differential geometry tangent space is $4mn$. Then $\dim N_{\mathcal{V}_{sr}} = 0$, and S_j is singular. It follows that $\mathcal{V}_{sr}^{ns} = \mathcal{M}_r$.

Next, we will prove (16) is satisfied. Choose a point $A = USV^T \in \mathcal{M}_r$, where $U \in \mathbb{R}^{4m \times 4r}$, $S = \text{diag} \{\Sigma, \Sigma, \Sigma, \Sigma\}$ and $V \in \mathbb{R}^{4n \times 4r}$. Denote

$$\mathcal{T}_i = \begin{pmatrix} \mathbb{R}^{r \times r} & \mathbb{R}^{r \times (n-r)} \\ \mathbb{R}^{(m-r) \times r} & 0 \end{pmatrix}, i = 0, 1, 2, 3.$$

By Lemma 2.2, the tangent space of \mathcal{M}_r at A can be expressed as

$$\mathcal{T}_{\mathcal{M}_r}(A) = \left\{ [U, U_\perp] \begin{pmatrix} \mathcal{T}_0 & -\mathcal{T}_1 & \mathcal{T}_2 & -\mathcal{T}_3 \\ \mathcal{T}_1 & \mathcal{T}_0 & -\mathcal{T}_3 & \mathcal{T}_2 \\ -\mathcal{T}_2 & \mathcal{T}_3 & \mathcal{T}_0 & -\mathcal{T}_1 \\ \mathcal{T}_3 & -\mathcal{T}_2 & \mathcal{T}_1 & \mathcal{T}_0 \end{pmatrix} [V, V_\perp]^T \right\}.$$

Then, it is easy to prove $T_{\mathcal{M}_*}(A) = \text{span}(W)$, with

$$W = \begin{pmatrix} 0 & -\tilde{\mathbf{1}}^{m \times n} & \tilde{\mathbf{1}}^{m \times n} & -\tilde{\mathbf{1}}^{m \times n} \\ \tilde{\mathbf{1}}^{m \times n} & 0 & -\tilde{\mathbf{1}}^{m \times n} & \tilde{\mathbf{1}}^{m \times n} \\ -\tilde{\mathbf{1}}^{m \times n} & \tilde{\mathbf{1}}^{m \times n} & 0 & -\tilde{\mathbf{1}}^{m \times n} \\ \tilde{\mathbf{1}}^{m \times n} & -\tilde{\mathbf{1}}^{m \times n} & \tilde{\mathbf{1}}^{m \times n} & 0 \end{pmatrix} \text{ with } \tilde{\mathbf{1}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}.$$

After that we can obtain (16). □

It follows from Theorem 2.11 that \mathcal{V} is an irreducible variety with dimension $m = 3(m+n)r - 3r^2$, then by Lemma 2.6, \mathcal{V}^{ns} can be chosen as a manifold with dimension $m = 3(m+n)r - 3r^2$. Moreover, the set of nonsingular points of \mathcal{V}_{sr} forms the manifold \mathcal{M}_r . Hence, by Lemma 2.7, every intersection point of $\mathcal{M}_r \cap \mathcal{M}_* \cap \mathcal{V}^{ns}$ is a non-trivial intersection point, i.e., the angle between \mathcal{M}_r and \mathcal{M}_* is well defined. Denote $\mathcal{V}^{ns, nt} \subset \mathcal{V}$ as the set of all points in $\mathcal{M}_r \cap \mathcal{M}_* \cap \mathcal{V}^{ns}$ that are nontangential with respect to the manifolds \mathcal{M}_r and \mathcal{M}_* . Then by (16) and Lemma 2.8, we have $\mathcal{V}^{ns, nt} \neq \emptyset$, which is saying that nontangentiality at one single intersection point implies nontangentiality at all the points of the manifold.

Based on the above results, we can get the main results of this section.

Theorem 2.13. *The set $\mathcal{V}^{ns, nt} = \mathcal{M}_r \cap \mathcal{M}_* = \mathcal{M}_{r*}$ is an $3(m+n)r - 3r^2$ dimensional manifold. Its complement $\mathcal{V} \setminus \mathcal{V}^{ns, nt}$ is a finite set of connected manifolds of lower dimension.*

3 Alternating Projections on Manifolds

In this section, the alternating projection method is chosen to solve the problem (2). The basic idea of alternating projections is to find a point in the intersection of two sets by

iteratively projecting a point into one set and then the other. Here, one manifold is the fixed rank r quaternion matrix set \mathcal{M}_r given as (5), and the other one is the pure quaternion matrix set \mathcal{M}_* given as (6). It follows that (2) can be rewritten as finding the nearest matrix in the set \mathcal{M}_{r*} given as (7), i.e., the intersection of the above two manifolds.

We first introduce two projections that project the given matrix onto the two matrix sets, respectively. Similar to the real and complex matrix cases, the Eckart-Young-Mirsky low-rank approximation theorem [11] still hold for quaternion matrices. With the singular value decomposition of quaternion matrix given in Theorem 2.1, the projection onto fixed rank matrix set \mathcal{M}_r can be expressed as

$$\pi_1(\mathbf{X}) = \sum_{i=1}^r \sigma_i(\mathbf{X}) \mathbf{u}_i(\mathbf{X}) \mathbf{v}_i^T(\mathbf{X}), \quad (17)$$

where $\sigma_i(\mathbf{X})$ are the r first singular values of \mathbf{X} , and $\mathbf{u}_i(\mathbf{X}), \mathbf{v}_i(\mathbf{X})$ are the first r columns of the unitary matrices of \mathbf{U} and \mathbf{V} given in Theorem 2.1, respectively. In addition, for an arbitrary quaternion matrix $\mathbf{X} = X_0 + X_1\mathbf{i} + X_2\mathbf{j} + X_3\mathbf{k} \in \mathbb{H}^{m \times n}$, the projection onto the affine manifold \mathcal{M}_* is exactly removing the real part, i.e.,

$$\pi_2(\mathbf{X}) = \pi_2(X_0 + X_1\mathbf{i} + X_2\mathbf{j} + X_3\mathbf{k}) = X_1\mathbf{i} + X_2\mathbf{j} + X_3\mathbf{k}. \quad (18)$$

Recall that \mathcal{M}_r and \mathcal{M}_* are fix rank manifold and affine manifold introduced in Section 2. Then the projection mappings may not be single valued. We write $\pi_1(\mathbf{X})$ and $\pi_2(\mathbf{X})$ to denote an arbitrarily closest point to \mathbf{X} on the manifolds \mathcal{M}_r and \mathcal{M}_* , respectively.

Nevertheless, the projection onto the intersection \mathcal{M}_{r*} cannot be computed efficiently. We use $\pi(\mathbf{X})$ to denote an arbitrarily closest point to \mathbf{X} on the intersection \mathcal{M}_{r*} . Furthermore, the convergence of the alternating projections cannot be guaranteed in general even when the two non-convex sets have a nonempty intersection, which is different from the convex case. For instance, suppose $\mathbf{K} = \mathbb{R}^2$ and denote $\mathcal{M}_1 = \{t, (t+1)(3-t)/4 : t \in \mathbb{R}\}$ and $\mathcal{M}_2 = \mathbb{R} \times \{0\}$. It easy to see that $\pi_1((1,0)) = (1,1)$ and $\pi_2((1,1)) = (1,0)$, then the sequence of alternating projections does not convergence. Therefore, it is more difficult to consider alternating projections on non-linear manifolds than convex sets.

The following algorithm describes the alternating projections method. of problem (2).

Algorithm 1 Alternating projections on manifolds

Input: Given a quaternion matrix $\mathbf{A} \in \mathbb{H}^{m \times n}$ this algorithm computes optimal rank- r pure quaternion matrix approximation.

- 1: Initialize $\mathbf{X}_0 = \mathbf{A}$;
- 2: **for** $k = 1, 2, \dots$
- 3: $\mathbf{Y}_{k+1} = \pi_1(\mathbf{X}_k)$;
- 4: $\mathbf{X}_{k+1} = \pi_2(\mathbf{Y}_{k+1})$;
- 5: **end**

Output: \mathbf{X}_k when the stopping criterion is satisfied.

Combining the above results with Theorem 2.13 and Theorem 5.1 in [3], we can obtain the main result of this paper.

Theorem 3.1. *Let \mathcal{M}_r , \mathcal{M}_* and $\mathcal{M}_{r*} = \mathcal{M}_r \cap \mathcal{M}_*$ be given as (5)-(7), respectively. The projections onto the manifolds \mathcal{M}_r and \mathcal{M}_* are given in (17) and (18), denote π as the projection onto the manifold \mathcal{M}_{r*} . Suppose that $\mathbf{B} \in \mathcal{M}_{r*}$ is a non-tangential intersection point of \mathcal{M}_r and \mathcal{M}_* , then for any given $\epsilon > 0$ and $1 > c > \sigma(\mathbf{B})$, there exist an $r > 0$*

such that for any $\mathbf{A} \in \mathcal{B}(\mathbf{B}, r)$ the sequence $\{\mathbf{X}_k\}_{k=0}^\infty$ generated by the alternating projections algorithm initializing from \mathbf{A} satisfies the following results:

- (1) the sequence converges to a point $\mathbf{X}_\infty \in \mathcal{M}_r \cap \mathcal{M}_*$,
- (2) $\|\mathbf{X}_\infty - \pi(\mathbf{A})\| \leq \epsilon \|\mathbf{A} - \pi(\mathbf{A})\|$,
- (3) $\|\mathbf{X}_\infty - \mathbf{X}_k\| \leq \text{const} \cdot c^k \|\mathbf{A} - \pi(\mathbf{A})\|$.

4 Initialization for Alternating Projections Algorithm

Theoretically, Theorem 3.1 implies that the alternating projections algorithm linearly converges to a good approximation to $\pi(\mathbf{A})$ assuming that \mathbf{A} is in some neighborhood $\mathcal{N}_{\epsilon, c}$ of the intersection manifold. Nevertheless, it is hard to check whether the original matrix is inside such a neighborhood or not since there is no explicit formula for the radius function $r_{\epsilon, c}$ in terms of the given scalars ϵ and c . Therefore, it is necessary to design an initialization strategy to make the alternating projections method more practical and reliable.

Recall the target projection is as follows

$$\begin{aligned} \pi(\mathbf{A}) \in \arg \min \quad & \|\mathbf{X} - \mathbf{A}\|_{\mathbb{F}}^2, \\ \text{s.t.} \quad & \text{Re}(\mathbf{X}) = 0, \\ & \text{rank}(\mathbf{X}) = r. \end{aligned} \quad (19)$$

We aim at an initial point X^0 which is close to the intersection of the two manifolds and $\pi(X^0) \approx \pi(A)$. If we apply a convergent iterative methods to the optimization problem (19), then it is reasonable to regard an iterate after several steps as such a good initial point. The reasons why we do not use this convergent iterative methods are (i) the convergence of this guaranteed method could be pretty slow and (ii) the computational cost of the alternating projections is generally much cheaper.

We reformulate the projection (19) to an unconstrained problem

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{A}\|_{\mathbb{F}}^2 + \delta_{\mathcal{S}_0}(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\|_{\mathbb{F}}^2 + \delta_{\mathcal{S}_r}(\mathbf{X}), \quad (20)$$

where δ_C denotes the indicator function of the set C . To adapt the convergence conditions which will be discussed shortly, we further relax the problem (5.4) to

$$\min \underbrace{\frac{1}{2} \|\mathbf{X} - \mathbf{A}\|_{\mathbb{F}}^2 + \frac{\tau}{2} \|\text{Re}(\mathbf{X})\|_{\mathbb{F}}^2}_{f(\mathbf{X})} + \underbrace{\frac{1}{2} \|\mathbf{X} - \mathbf{A}\|_{\mathbb{F}}^2 + \delta_{\mathcal{S}_r}(\mathbf{X})}_{g(\mathbf{X})}. \quad (21)$$

Note that the problems (5.4) and (5.4) are equivalent when τ approaches the infinity.

Li and Pong [25] proposed and investigate the Douglas-Rachford splitting method (DRSM) for solving the nonconvex optimization problem

$$\min f(x) + g(x).$$

The DRSM iterates

$$\begin{cases} y^{k+1} \in \text{prox}_{\alpha f}(x^k) := \arg \min_y \{f(y) + \frac{1}{2\alpha} \|y - x^k\|_2^2\}, \\ z^{k+1} \in \text{prox}_{\alpha g}(2y^{k+1} - x^k) := \arg \min_z \{g(z) + \frac{1}{2\alpha} \|2y^{k+1} - x^k - z\|_2^2\}, \\ x^{k+1} = x^k + (z^{k+1} - y^{k+1}), \end{cases}$$

Assuming the existence of a cluster point, they proved the following conditions can guarantee the global convergence to a stationary point:

- (1) f has a Lipschitz continuous gradient whose Lipschitz continuity modulus is bounded by L ,
- (2) g is a proper closed function,
- (3) f and g are semi-algebraic functions,
- (4) $0 < \alpha < \frac{1}{L}$.

In the problem (5.4), the function $f(\mathbf{X})$ is quadratic and thus it is semi-algebraic and has Lipschitz continuous gradient whose Lipschitz constant is $1 + \tau$. The manifold \mathcal{S}_r can be characterized by

$$\{\mathbf{Y} : \det(\mathbf{Y}_{r+1}) = 0 \text{ for any } (r+1)\text{-by-}(r+1) \text{ submatrix } \mathbf{Y}_{r+1}\}.$$

Hence, this set is a semi-algebraic set, which implies that its indicator function is a semi-algebraic function. That is, $g(\mathbf{X})$ is also a semi-algebraic function. Therefore, the global convergence to a stationary point can be guaranteed as long as we choose the stepsize α less than $\frac{1}{1+\tau}$.

Furthermore, the proximal operator for αf is implemented by

$$\begin{aligned} \text{prox}_{\alpha f}(\mathbf{Y}) = & \left(\frac{\alpha}{1+\alpha+\alpha\tau} A_1 + \frac{1}{1+\alpha+\alpha\tau} Y_1 \right) + \left(\frac{\alpha}{1+\alpha} A_2 + \frac{1}{1+\alpha} Y_2 \right) \mathbf{i} \\ & + \left(\frac{\alpha}{1+\alpha} A_3 + \frac{1}{1+\alpha} Y_3 \right) \mathbf{j} + \left(\frac{\alpha}{1+\alpha} A_4 + \frac{1}{1+\alpha} Y_4 \right) \mathbf{k}, \end{aligned}$$

and the proximal operator for αg is

$$\text{prox}_{\alpha g}(\mathbf{Y}) = \pi_1 \left(\frac{\alpha}{1+\alpha} \mathbf{A} + \frac{1}{1+\alpha} \mathbf{Y} \right),$$

i.e., the truncation to a rank- r quaternion matrix. To sum up, the DRSM for solving (5.4) iterates

$$\begin{cases} \mathbf{Y}^{k+1} = \left(\frac{\alpha}{1+\alpha+\alpha\tau} A_1 + \frac{1}{1+\alpha+\alpha\tau} X_1^k \right) + \left(\frac{\alpha}{1+\alpha} A_2 + \frac{1}{1+\alpha} X_2^k \right) \mathbf{i} \\ \quad + \left(\frac{\alpha}{1+\alpha} A_3 + \frac{1}{1+\alpha} X_3^k \right) \mathbf{j} + \left(\frac{\alpha}{1+\alpha} A_4 + \frac{1}{1+\alpha} X_4^k \right) \mathbf{k}, \\ \mathbf{Z}^{k+1} \in \pi_1 \left(\frac{\alpha}{1+\alpha} \mathbf{A} + \frac{2}{1+\alpha} \mathbf{Y}^{k+1} - \frac{1}{1+\alpha} \mathbf{X}^k \right), \\ \mathbf{X}^{k+1} = \mathbf{X}^k + (\mathbf{Z}^{k+1} - \mathbf{Y}^{k+1}). \end{cases} \quad (22)$$

We perform the DRSM for a fixed number steps and then apply the generated iterate \mathbf{Y}^k as the initial point of the alternating projections method. Note that this initialization strategy is still heuristic and the improvement using this method will be numerically illustrated in the following section.

5 Numerical Experiments

In this paper, we focus on searching an optimal rank- r pure quaternion matrix approximation of a given quaternion matrix. Although this problem is difficult, there exist some suboptimal methods to solve it. For example, in [17], the authors do the rank r truncation of a given quaternion matrix and then take three imaginary parts as an approximation of the given quaternion matrix. This method is called “QsvdTr” in the sequel. However, if the real part of the quaternion matrix is removed, the rank of the quaternion matrix generally changes. This fact can be guaranteed by the following proposition (The proof can be found in the Appendix).

Proposition 5.1. For an arbitrary quaternion $\mathbf{A} = A_0 + A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k} \in \mathbb{H}^{m \times n}$, with $\text{rank}(\mathbf{A}) = r \leq \min\{m, n\}/4$, we denote the pure quaternion part of \mathbf{A} as $\mathbf{A}_p = A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k}$, then

$$r \leq \text{rank}(\mathbf{A}_p) \leq 4r.$$

In Subsection 5.1-5.3, the performances of the “QsvdTr” algorithm given in [17] and the “AltProj” algorithm proposed in Algorithm 2 are compared by testing synthetic data, random data and color images, respectively. We use the running time and the objective function values, i.e., $\|\mathbf{X}^k - \mathbf{A}\|_F$ to compare the results derived by “QsvdTr” and “AltProj”. All the experiments are performed under Windows 10 and MATLAB R2018a running on a desktop (Intel Core i7, @ 3.40GHz, 8.00G RAM).

5.1 Synthetic Data

In our first example, we compare the “AltProj” algorithm and “QsvdTr” algorithm by finding the optimal rank 4 pure quaternion approximation of the pure quaternion matrix

$$\mathbf{A} = \begin{pmatrix} 0.37 & -0.79 & 0.04 & -0.73 & -0.06 \\ -1.42 & -0.10 & 1.01 & 1.59 & -1.59 \\ -0.34 & 0.38 & 1.30 & -0.66 & 1.08 \\ -1.98 & 0.83 & 0.22 & -0.77 & 0.70 \\ -0.38 & -0.14 & 0.86 & 0.54 & 1.65 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0.29 & -0.38 & -0.13 & -1.77 & 0.20 \\ 0.70 & -0.69 & 0.83 & -0.16 & -0.52 \\ -1.15 & 1.00 & -1.97 & 0.63 & 1.57 \\ 1.86 & -1.14 & 0.12 & -1.27 & 0.77 \\ 2.37 & 0.15 & 0.26 & -0.30 & -0.59 \end{pmatrix} \mathbf{j} \\ + \begin{pmatrix} 0.33 & 0.74 & -1.40 & -0.77 & 0.86 \\ 1.13 & -1.32 & 0.36 & -0.02 & 0.50 \\ 0.25 & -0.68 & 0.36 & -0.71 & 0.77 \\ 0.56 & -0.35 & 0.92 & 0.87 & -0.58 \\ 0.64 & -1.59 & 0.37 & -1.51 & 0.19 \end{pmatrix} \mathbf{k}.$$

By applying Algorithm 1 on \mathbf{A} , we can get a rank 4 pure quaternion approximation as

$$\mathbf{A}_4 = \begin{pmatrix} 0.50 & -0.73 & -0.01 & -0.68 & 0.08 \\ -1.34 & -0.11 & 1.05 & 1.69 & -1.50 \\ -0.29 & 0.36 & 1.24 & -0.64 & 1.11 \\ -2.01 & 0.81 & 0.11 & -0.83 & 0.56 \\ -0.37 & -1.12 & 0.94 & 0.57 & 1.70 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0.23 & -0.50 & 0.08 & -1.71 & 0.37 \\ 0.65 & -0.95 & 0.76 & -0.17 & -0.47 \\ -1.17 & 0.91 & -1.96 & 0.62 & 1.63 \\ 1.92 & -1.01 & 0.05 & -1.30 & 0.74 \\ 2.35 & 0.09 & 0.28 & -0.28 & -0.58 \end{pmatrix} \mathbf{j} \\ + \begin{pmatrix} 0.39 & 0.65 & -1.41 & -0.66 & 0.92 \\ 1.07 & -1.34 & 0.34 & -0.04 & 0.36 \\ 0.22 & -0.76 & 0.37 & -0.70 & 0.77 \\ 0.51 & -0.44 & 0.92 & 0.82 & -0.54 \\ 0.67 & -1.53 & 0.35 & -1.50 & 0.16 \end{pmatrix} \mathbf{k}.$$

The singular value decomposition of \mathbf{A}_4 can be expressed as $\mathbf{A}_4 = \mathbf{U}\mathbf{S}\mathbf{V}$ with

$$\mathbf{U} = \begin{pmatrix} -0.02 & -0.11 & 0.02 & 0.17 & -0.43 \\ -0.10 & -0.24 & -0.04 & -0.19 & 0.07 \\ 0.21 & 0.14 & -0.34 & 0.17 & 0.02 \\ 0.06 & 0.24 & 0.16 & 0.30 & 0.36 \\ 0.04 & -0.01 & 0.10 & 0.04 & -0.20 \end{pmatrix} + \begin{pmatrix} -0.27 & -0.14 & 0.30 & 0.25 & 0.25 \\ 0.35 & 0.29 & 0.14 & -0.37 & 0.28 \\ -0.05 & -0.50 & -0.16 & -0.05 & 0.18 \\ 0.33 & -0.30 & 0.18 & -0.05 & -0.11 \\ 0.08 & -0.05 & 0.12 & 0.34 & -0.03 \end{pmatrix} \mathbf{i} \\ + \begin{pmatrix} -0.15 & -0.05 & 0.13 & -0.10 & 0.16 \\ -0.19 & -0.12 & -0.10 & -0.14 & 0.46 \\ 0.22 & -0.04 & 0.11 & -0.22 & 0.17 \\ -0.40 & 0.02 & -0.32 & -0.25 & -0.08 \\ -0.32 & 0.53 & -0.22 & 0.06 & 0.05 \end{pmatrix} \mathbf{j} + \begin{pmatrix} -0.05 & -0.32 & -0.34 & 0.18 & 0.37 \\ -0.31 & -0.07 & 0.08 & 0.17 & -0.17 \\ -0.23 & 0.01 & 0.49 & 0.21 & 0.07 \\ 0.00 & 0.03 & -0.23 & 0.23 & 0.07 \\ -0.33 & 0.11 & 0.24 & -0.44 & -0.09 \end{pmatrix} \mathbf{k}, \\ \mathbf{V} = \begin{pmatrix} 0.65 & 0.31 & -0.54 & 0.30 & 0.32 \\ 0.39 & -0.08 & -0.24 & 0.18 & 0.31 \\ -0.09 & 0.05 & 0.10 & 0.18 & 0.31 \\ 0.39 & 0.28 & -0.10 & 0.24 & 0.13 \\ -0.06 & -0.43 & 0.12 & 0.41 & 0.10 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -0.12 & 0.12 & 0.06 & -0.48 & 0.18 \\ 0.03 & -0.06 & -0.07 & -0.16 & 0.38 \\ 0.05 & 0.41 & 0.42 & 0.27 & 0.16 \\ -0.11 & -0.14 & 0.15 & 0.11 & 0.04 \end{pmatrix} \mathbf{i} \\ + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -0.19 & -0.07 & 0.04 & 0.05 & -0.29 \\ 0.38 & 0.07 & -0.59 & 0.10 & -0.28 \\ 0.14 & 0.04 & -0.02 & 0.18 & 0.03 \\ -0.06 & -0.19 & -0.08 & 0.06 & -0.13 \end{pmatrix} \mathbf{j} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.11 & 0.07 & -0.03 & -0.28 & 0.37 \\ -0.13 & -0.26 & -0.22 & -0.22 & -0.18 \\ 0.00 & -0.07 & 0.07 & 0.39 & -0.16 \\ -0.04 & 0.55 & 0.08 & -0.04 & -0.43 \end{pmatrix} \mathbf{k},$$

and

$$S = \begin{pmatrix} 5.90 & 0 & 0 & 0 & 0 \\ 0 & 4.47 & 0 & 0 & 0 \\ 0 & 0 & 2.96 & 0 & 0 \\ 0 & 0 & 0 & 1.75 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Moreover, by Theorem 2.1 the optimal rank 4 approximation of \mathbf{A} can be expressed as $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_{re} + \tilde{\mathbf{A}}_p$, where

$$\begin{aligned} \tilde{\mathbf{A}}_{re} &= \begin{pmatrix} 0.03 & 0.17 & 0.10 & 0.03 & -0.02 \\ -0.05 & -0.03 & 0.14 & 0.03 & 0.06 \\ 0.01 & 0.02 & 0.06 & 0.03 & 0.04 \\ 0.03 & -0.07 & -0.07 & 0.02 & 0.03 \\ -0.03 & 0.02 & 0.03 & -0.02 & -0.02 \end{pmatrix}, \\ \tilde{\mathbf{A}}_p &= \begin{pmatrix} 0.46 & -0.75 & 0.03 & -0.70 & 0.04 \\ 0.06 & -1.36 & -0.10 & 1.66 & -1.53 \\ -0.04 & -0.30 & 0.37 & 1.26 & -0.65 \\ 0.03 & -2.00 & 0.82 & 0.14 & -0.82 \\ -0.02 & -0.37 & -0.13 & 0.91 & 0.56 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0.25 & -0.47 & 0.01 & -1.73 & 0.31 \\ 0.67 & -0.86 & 0.78 & -0.16 & -0.48 \\ -1.17 & 0.95 & -1.96 & 0.63 & 1.62 \\ 1.91 & -1.05 & -0.08 & -1.29 & 0.75 \\ 2.35 & 0.11 & 0.27 & -0.29 & -0.58 \end{pmatrix} \mathbf{j} \\ &+ \begin{pmatrix} 0.38 & 0.69 & -1.42 & -0.70 & 0.89 \\ 1.09 & -1.34 & 0.34 & -0.03 & 0.40 \\ 0.23 & -0.73 & 0.36 & -0.71 & 0.77 \\ 0.53 & -0.41 & 0.92 & 0.84 & -0.55 \\ 0.66 & -1.55 & 0.36 & -1.51 & 0.17 \end{pmatrix} \mathbf{k}. \end{aligned}$$

In this case, $\tilde{\mathbf{A}}_p$ is often chosen as the optimal pure quaternion rank 4 approximation. However, $\text{rank}(\tilde{\mathbf{A}}_p) = 5$, i.e., $\tilde{\mathbf{A}}_p$ is essentially not a rank 4 approximation. Recall Proposition 5.1 that the rank of a quaternion matrix will increase if its real part is removed. Then if one want to find a rank r pure quaternion approximation, less than r truncation of \mathbf{A} is suitable. So, we can get a rank 4 pure quaternion approximation

$$\begin{aligned} \mathbf{A}_p &= \begin{pmatrix} 1.05 & -0.78 & 0.37 & -0.59 & 0.13 \\ -1.32 & 0.25 & 0.69 & 1.07 & -0.26 \\ 0.20 & -0.06 & 0.31 & 0.03 & -0.01 \\ -1.30 & 0.58 & 0.16 & 0.73 & 0.02 \\ -0.27 & -0.38 & 0.98 & 0.42 & -0.03 \end{pmatrix} \mathbf{i} + \begin{pmatrix} 0.56 & -0.14 & -0.15 & -0.32 & 0.00 \\ 0.68 & -0.44 & 0.54 & -0.42 & 0.33 \\ -0.78 & -0.93 & -0.65 & 0.24 & 0.12 \\ 1.52 & -1.07 & 0.39 & -0.97 & 0.28 \\ 1.28 & -0.47 & 0.08 & -0.88 & 0.37 \end{pmatrix} \mathbf{j} \\ &+ \begin{pmatrix} 0.17 & 0.09 & -0.57 & -0.30 & 0.04 \\ 1.21 & -1.20 & 0.86 & -0.40 & -0.14 \\ 0.87 & -0.42 & 0.17 & -0.61 & 0.30 \\ -0.05 & -0.72 & 0.84 & 0.40 & -0.36 \\ 1.26 & -1.14 & 0.41 & -0.63 & -0.08 \end{pmatrix} \mathbf{k}, \end{aligned}$$

by deleting the real part of its optimal rank 1 approximation of \mathbf{A} . In this case, the objective function values are $\|\mathbf{A} - \mathbf{A}_4\|_F = 0.6479$ and $\|\mathbf{A} - \mathbf{A}_p\|_F = 5.3579$, respectively. We see that approximation derived by the proposed ‘AltProj’ algorithm is better than that derived by the ‘QsvdTr’ algorithm.

5.2 Random Matrices

In our second experiment, we use random low rank pure quaternion matrix to illustrate the validity of Algorithm 2. In the proposed ‘AltProj’ algorithm, the maximum number of iterations is chosen as 5000 steps. And the iterations stops when the residual, i.e., the Frobenius norms of the real part is less than 10^{-6} . Since it is hard to generate a random low rank pure quaternion matrix directly, the following alternative method is applied. We first generate m -by- n quaternion matrices $\mathbf{A} = A_0 + A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k}$, where the matrix entries of $A_i, i = 0, 1, 2, 3$ follow the standard normal distribution. Random quaternion matrices with ranks 1, 2, 3, 4, 5 and 10 can be derived by applying QSVD truncations on \mathbf{A} , respectively. It follows Proposition 5.1 that the rank of quaternion matrix will increase when the real part is

deleted. In addition, when the quaternion matrix is randomly generated with small rank, the columns of the different full rank decompositions with three kinds of conjugate definitions are always independent then $\text{rank}(\mathbf{A}_p) = 4\text{rank}(\mathbf{A})$ holds at most times. Although this result cannot be guaranteed to be correct in general, we often use this method to obtain approximate solutions of some problems in practical application. Then we can get ranks 4, 8, 12, 16, 20 and 40 pure quaternion matrices by setting the real parts of the these low rank quaternion matrices to be zeros, respectively.

Tables 1 shows the running times and the objective function values of the computed solutions from the proposed “AltProj” algorithm and the “QsvdTr” algorithm for ranks 4, 8, 12, 16, 20 and 40 random pure quaternion matrices sets of sizes 100-by-100, 200-by-200 and 500-by-500, respectively. When the input quaternion matrix \mathbf{A} is exactly a low rank pure quaternion matrix, the proposed “AltProj” algorithm can provide exact recovery results in the first iteration. However, there is no guarantee that the “QsvdTr” algorithm can determine the low rank pure quaternion matrix. In the tables, it is clear that the “QsvdTr” algorithm cannot obtain the underlying low rank factorization. The running times of the “AltProj” algorithm is nearly two times of “QsvdTr” algorithm.

Table 1: The running times (Time) and the objective function values (OBF-value) by different algorithms for low rank pure quaternion approximations of random low rank pure quaternion matrices.

method	item	100-by-100 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	8.89	12.78	15.61	18.05	20.22	28.03
	Time (s)	0.28	0.28	0.29	0.29	0.29	0.30
AltProj	OBF-value	4.09e-14	1.65e-13	6.80e-13	1.90e-13	1.30e-13	1.87e-13
	Ttime (s)	0.56	0.57	0.57	0.58	0.57	0.58
method	item	200-by-200 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	12.58	17.79	21.91	25.31	28.24	40.22
	Time (s)	1.40	1.42	1.41	1.42	1.46	1.44
AltProj	OBF-value	6.07e-14	1.28e-13	2.86e-13	3.40e-13	2.25e-13	3.21e-13
	Time (s)	2.79	2.80	2.83	2.84	2.83	2.85
method	item	500-by-500 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	19.57	27.61	33.98	39.07	43.92	62.39
	Time (s)	19.11	18.94	19.03	19.13	19.11	19.27
AltProj	OBF-value	9.50e-14	2.70e-13	3.19e-13	6.13e-13	4.66e-13	6.57e-13
	Time (s)	38.06	38.14	38.45	38.48	38.21	38.48

In our third experiment, we use random pure quaternion matrix to compare the two algorithms, where the low rank minimizer is unknown in this setting. The maximum number of iterations and the tolerance of the residual relate to the “AltProj” algorithm are chosen as 5000 steps and 10^{-6} , respectively. We randomly generate m -by- n pure quaternion matrices $\mathbf{A} = A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k}$, where the matrix entries of $A_i, i = 1, 2, 3$ follow the standard normal distribution. Then by applying the “QsvdTr” algorithm and the proposed “AltProj” algorithm to \mathbf{A} , we can find its optimal ranks 4, 8, 12, 16, 20 and 40 pure quaternion approximations, respectively. Table 2 shows that the running times and the objective function values of the computed solution \mathbf{X}^k from the proposed “AltProj” algorithm and

the “QsvdTr” algorithm. We see from Table 2 that the objective function values computed by the proposed “AltProj” algorithm are smaller than that derived by the testing “QsvdTr” algorithm, although the corresponding running times are longer.

Table 2: The running times (Time) and the objective function values (OBF-value) by different algorithms for low rank pure quaternion approximations of random pure quaternion matrices.

method	item	100-by-100 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	171.00	166.97	163.10	159.43	155.78	139.52
	Time (s)	0.32	0.33	0.32	0.31	0.32	0.31
AltProj	OBF-value	168.48	160.66	152.29	143.66	1.35.24	91.12
	Time (s)	1.73	33.36	2.83	48.37	24.01	12.70
method	item	200-by-200 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	342.33	338.21	334.31	330.57	326.79	308.86
	Time (s)	1.56	1.55	1.51	1.55	1.56	1.55
AltProj	OBF-value	339.88	332.94	325.36	317.80	310.34	268.59
	Time (s)	13.42	11.71	11.31	14.60	29.84	30.48
method	item	500-by-500 quaternion matrix					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	861.25	857.10	853.01	848.98	844.98	825.60
	Time (s)	20.58	20.67	20.32	20.28	20.48	21.75
AltProj	OBF-value	860.02	853.28	846.58	839.92	823.73	794.54
	Time (s)	72.04	184.56	175.56	1460.45	405.81	1099.45

5.3 Color Images

In this subsection, we employ the color images ‘peppafamily’, ‘pepper’ and ‘colortexture’ with sizes 200-by-200 to compare “QsvdTr” algorithm and “AltProj” algorithm in terms of objective function values and time. The maximum number of iterations and the tolerance of the residual relate to the “AltProj” algorithm are chosen as 5000 steps and 10^{-6} , respectively. The original three color images employed in this subsection are shown in the first column of Figure 1, and the rank 16 and 20 pure quaternion approximations by the “QsvdTr” algorithm are listed in the second and fourth columns, the optimal rank 16 and 20 pure quaternion approximations derived by our algorithm are listed in the third and fifth columns. For the two cases, we can see respectively that the images derived by the “AltProj” algorithm are better than those derived by “QsvdTr” algorithm in terms of visual quality. The “AltProj” algorithm can preserve more details than “QsvdTr” algorithm for the three testing images.

Moreover, we also compute the objective function values of ranks 4, 8, 12, 16, 20 and 40 approximations, respectively, which illustrates the validity of our method. The results and shown in Table 3. It can be seen that the objective function values obtained by “AltProj” algorithm are much lower than those by “QsvdTr” algorithm. For the “colortexture” and “pepper” images, the objective function values of the “AltProj” algorithm are nearly half of that derived by “QsvdTr” algorithm.

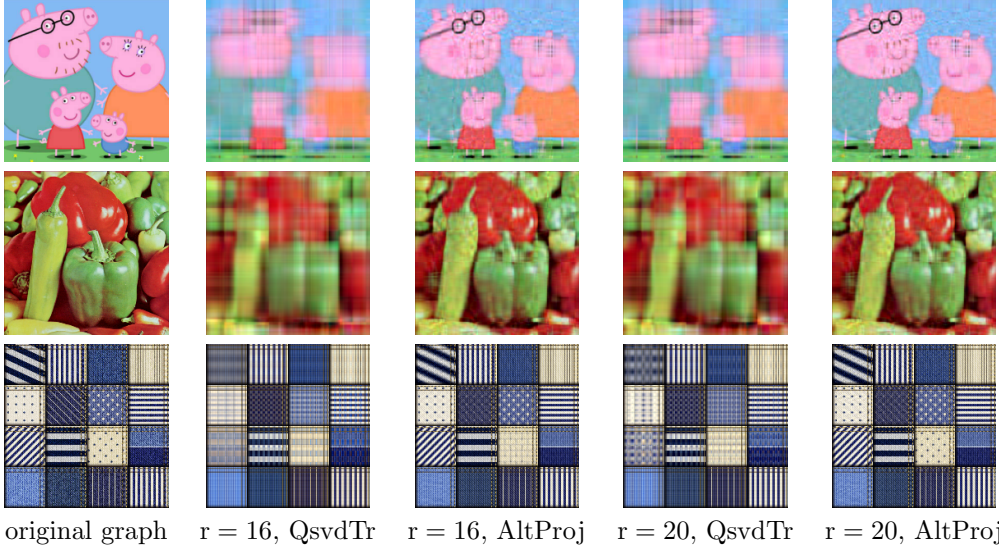


Figure 1: ranks 16 and 20 approximations of “peppafamily”, “pepper” and “colortexture” by QsvdTr and AltProj algorithms, respectively

Table 3: The running times (Time) and the objective function values (OBF-value) by different algorithms for low rank approximations of color images

method	item	peppafamily					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	13731.6	10944.1	9450.5	8742.4	8143.7	6470.5
	Time (s)	1.55	1.55	1.51	1.55	1.56	1.55
AltProj	OBF-value	9077.2	7351.9	6218.8	5439.9	4770.3	2726.0
	Time (s)	189.78	189.06	191.38	190.19	195.46	211.15
method	item	pepper					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	17831.5	14407.2	12587.8	11147.5	10153.1	7185.0
	Time (s)	1.67	1.57	1.68	1.56	1.59	1.56
AltProj	OBF-value	11392.7	8494.4	6804.4	5667.4	4949.5	2923.9
	Time (s)	201.76	206.77	221.51	206.21	193.88	199.59
method	item	colortexture					
		r=4	r=8	r=12	r=16	r=20	r=40
QsvdTr	OBF-value	21964.6	18261.4	15485.2	14034.7	13155.7	9530.4
	Time (s)	1.63	1.57	1.57	1.63	1.85	1.55
AltProj	OBF-value	14039.5	10864.6	8583.1	7199.8	6527.4	4496.8
	Time (s)	219.47	199.95	215.61	205.76	197.87	203.71

5.4 Initialization

The experiments in this section are conducted under Windows 10 and Matlab R2017a running on a desktop (Intel Core i7-8700, CPU @ 3.20GHz, 16.0G RAM). We still employ the three color images “peppafamily”, “pepper”, and “colortexture” to verify the effectiveness

of our proposed initialization strategy using the Douglas-Rachford splitting method in Section 4. We test the performances for (1) the alternating projections, (2) DRSM, (3) the alternating projections initialized by DRSM on the best rank-20 approximations to these three images. In the DRSM, we set the parameters τ and α adaptively. A large penalty parameter τ implies a good approximation to the original problem. However, it also leads to a small stepsize, which may cause a slow convergence and also overflows in the floating point arithmetic. Hence, we gradually increase τ but keep it constant when it is large enough:

$$\tau_0 = 1, \quad \tau_k = \begin{cases} 2\tau_{k-1}, & k \leq 1000, \\ \tau_{k-1}, & k > 1000. \end{cases}$$

We adopt the strategy for adaptively choosing the stepsize α in [25]:

$$\alpha_0 = \frac{150}{1 + \tau_0}, \quad \alpha_k = \max \left\{ 0.7\alpha_{k-1}, \frac{0.99}{1 + \tau_k} \right\},$$

which satisfy the convergence conditions when k is sufficiently large. When τ_k is large, the corresponding α_k will be close to zero. Then the coefficients in the iteration (22) will be close to either zero or $O(1)$ constants. Hence, a large parameter τ_k will not lead to numerical problems when iterating.

For the images “peppafamily” and “pepper”, we set the maximum total number of iterations as 5000 steps, where 500 steps of DRSM for initialization. For the image “colortexture”, for which the involved algorithms converge slower than the other images, we set the maximum total number of iterations as 10000 steps, where 3000 steps of DRSM for initialization. We also cease the iterations when the residual, the Frobenius norms of the real part, is less than 10^{-6} .

Figure 2 displays the residuals decreasing with the iterations and the running time. The convergence of the DRSM can be very slow although guaranteed. One can conclude from the numerical comparisons that the alternating projections converges much faster with the initialization by DRSM than simply taking the original image as the initial point.

We also present the singular values of the original images (quaternion matrices) and the results for different algorithms in Figure 3. The truncations of the singular values with our proposed initialization strategy are the most “clear” among these three algorithms. That is, the final results with the initialization are the closest to an actual rank-20 quaternion matrix.

6 Conclusion

We propose the alternating projections method for computing the optimal rank- r pure quaternion approximation to any pure quaternion matrix, which consists of the alternating projections onto the rank- r quaternion matrix manifold and the pure quaternion matrix manifold. The linear local convergence for the alternating projections method is proved employing the manifold structures. In order to guarantee the quality of the limit point and pursue a faster convergence rate, we also propose an initialization strategy using the Douglas-Rachford splitting method to search for an initial point in some neighborhood of the intersection manifold. Furthermore, we also conduct numerical experiments on both random matrices and real-world color images to illustrate the effect of our proposed alternating projections method and the initialization strategy.

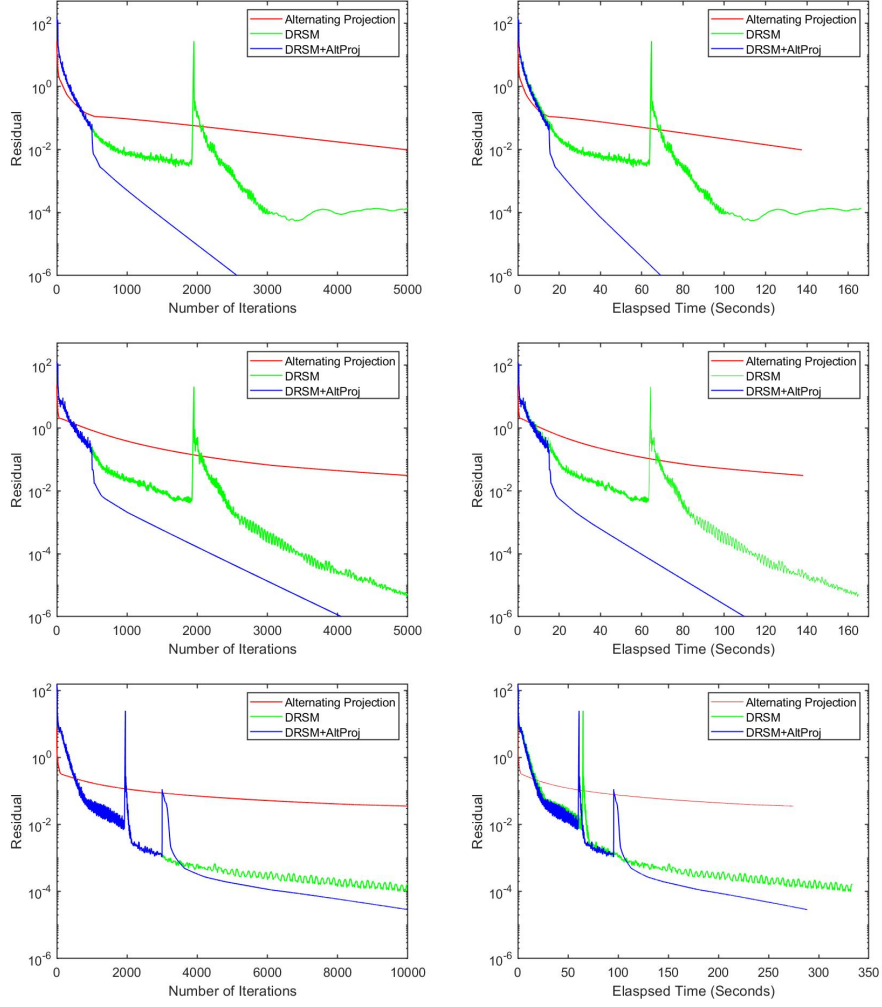


Figure 2: The residuals for the alternating projections, DRSM, the alternating projections initialized by DRSM. These three rows display the cases for “peppafamily”, “pepper”, and “colortexture”, respectively. The figures in the left column show the residuals against the number of iterations, and the figures in the right column show the residuals against the elapsed time.

References

- [1] P.A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [2] V.M. ADAMYAN AND D.Z. AROV, *A general solution of a problem in linear prediction of stationary processes*, Theory of Probability & Its Applications, 13 (1968), pp. 394-407.
- [3] F. ANDERSSON AND M. CARLSSON, *Alternating projections on nontangential mani-*

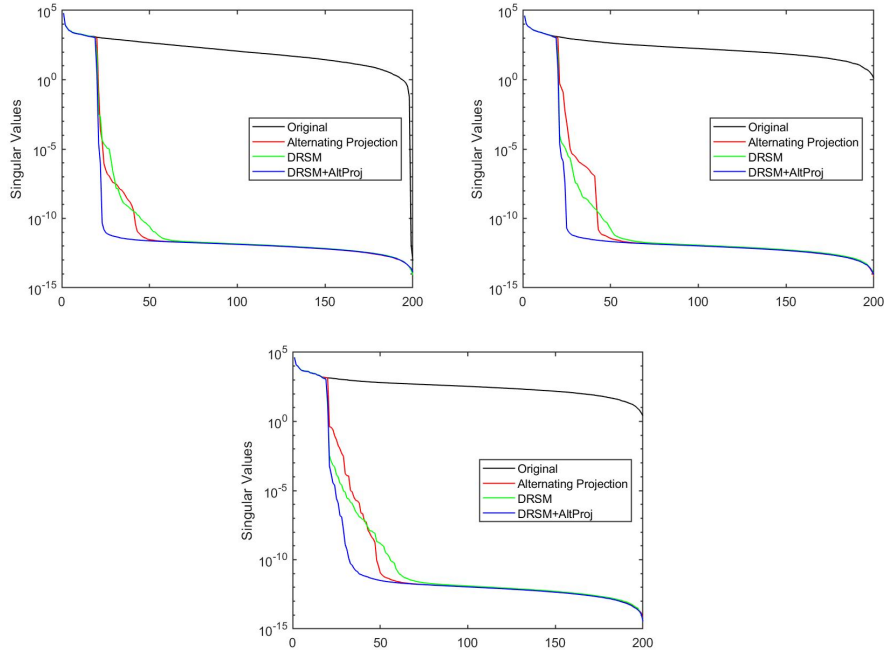


Figure 3: The singular values of the original images and the results for the alternating projections, DRSM, the alternating projections initialized by DRSM. These three figures display the cases for “peppafamily”, “pepper”, and “colortexture”, respectively.

folds, Constructive approximation, 38 (2013), pp. 489-525.

- [4] Q. BARTHÉLEMY, A. LARUE, AND J.I. MARS, *Color sparse representations for image processing: review, models, and prospects*, IEEE Transactions on Image Processing, 24 (2015), pp. 3978-3989.
- [5] H.H. BAUSCHKE AND J.M. BORWEIN, *On projection algorithms for solving convex feasibility problems*, SIAM review, 38 (1996), pp. 367-426.
- [6] J.A. CADZOW, *Signal enhancement-a composite property mapping algorithm*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 36 (1988), pp. 49-62.
- [7] J.H. CHANG, J.J. DING, ET AL., *Quaternion matrix singular value decomposition and its applications for color image processing*, in Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429), vol. 1, IEEE, 2003, pp. I-805.
- [8] P.L. COMBETTES, *Signal recovery by best feasible approximation*, IEEE transactions on Image Processing, 2 (1993), pp. 269-271.
- [9] T.A. ELL AND S.J. SANGWINE, *Hypercomplex fourier transforms of color images*, IEEE Transactions on image processing, 16 (2006), pp. 22-35.
- [10] P. FLETCHER AND S.J. SANGWINE, *The development of the quaternion wavelet transform*, Signal Processing, 136 (2017), pp. 2-15.

- [11] G.H. GOLUB AND C.F. VAN LOAN, *Matrix computations*, vol. 3, JHU press, 2012.
- [12] K.M. GRIGORIADIS, A.E. FRAZHO, AND R.E. SKELTON, *Application of alternating convex projection methods for computation of positive toeplitz matrices*, IEEE transactions on signal processing, 42 (1994), pp. 1873-1875.
- [13] C. HAMAKER AND D. SOLMON, *The angles between the null spaces of x rays*, Journal of mathematical analysis and applications, 62 (1978), pp. 1-23.
- [14] X. HAN, J. WU, L. YAN, L. SENHADJI, AND H. SHU, *Color image recovery via quaternion matrix completion*, in 2013 6th International Congress on Image and Signal Processing (CISP), vol. 1, IEEE, 2013, pp. 358-362.
- [15] N.J. HIGHAM, *Computing the nearest correlation matrix: a problem from finance*, IMA journal of Numerical Analysis, 22 (2002), pp. 329-343.
- [16] Z.G. JIA, M.K. NG, AND G.J. SONG, *Robust quaternion matrix completion with applications to image inpainting*, Numerical Linear Algebra with Applications, (2019), p. e2245.
- [17] Z.G. JIA, M. WEI, AND S.T. LING, *A new structure-preserving method for quaternion hermitian eigenvalue problems*, Journal of Computational and Applied Mathematics, 239 (2013), pp. 12-24.
- [18] Z.G. JIA, S.T. LING, AND M.X. ZHAO, *Color two-dimensional principal component analysis for face recognition based on quaternion model*, in International Conference on Intelligent Computing, Springer, 2017, pp. 177-189.
- [19] S. KAYALAR AND H.L. WEINERT, *Error bounds for the method of alternating projections*, Mathematics of Control, Signals and Systems, 1 (1988), pp. 43-59.
- [20] N.LE. BIHAN AND J. MARS, *Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing*, Signal processing, 84 (2004), pp. 1177-1199.
- [21] J.M. Lee, *Introduction to smooth manifolds*, Springer, 2013.
- [22] S. LEE, P.S. CHO, R.J. MARKS II, AND S. OH, *Conformal radiotherapy computation by the method of alternating projections onto convex sets*, Physics in Medicine & Biology, 42 (1997), p. 1065.
- [23] A. LEVI AND H. STARK, *Signal restoration from phase by projections onto convex sets*, JOSA, 73 (1983), pp. 810-822.
- [24] A. S. LEWIS AND J. MALICK, *Alternating projections on manifolds*, Mathematics of Operations Research, 33 (2008), pp. 216-234.
- [25] G. LI AND T.K. PONG, *Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems*, Math. Program., 159 (2016), pp. 371-401.
- [26] Y. LI, M. WEI, F. ZHANG, AND J. ZHAO, *A fast structure-preserving method for computing the singular value decomposition of quaternion matrices*, Applied Mathematics and Computation, 235 (2014), pp. 157-167.
- [27] ———, *Real structure-preserving algorithms of householder based transformations for quaternion matrices*, Journal of Computational and Applied Mathematics, 305 (2016), pp. 82-91.

- [28] S.C. PEI, J.J. DING, AND J.H. CHANG, *Efficient implementation of quaternion fourier transform, convolution, and correlation by 2-d complex fft*, IEEE Transactions on Signal Processing, 49 (2001), pp. 2783-2797.
- [29] S.J. SANGWINE, *Fourier transforms of colour images using quaternion or hypercomplex numbers*, Electronics letters, 32 (1996), pp. 1979-1980.
- [30] U. SHALIT, D. WEINSHALL, AND G. CHECHIK, *Online learning in the manifold of low-rank matrices*, in Advances in neural information processing systems, 2010, pp. 2128-2136.
- [31] ———, *Online learning in the embedded manifold of low-rank matrices*, Journal of Machine Learning Research, 13 (2012), pp. 429-458.
- [32] Ö. N. SUBAKAN AND B. C. VEMURI, *A quaternion framework for color image smoothing and segmentation*, International Journal of Computer Vision, 91 (2011), pp. 233-250.
- [33] K. TANABE, *Projection method for solving a singular system of linear equations and its applications*, Numerische Mathematik, 17 (1971), pp. 203-214.
- [34] B. VANDEREYCKEN, *Low-rank matrix completion by Riemannian optimization*, SIAM Journal on Optimization, 23 (2013), pp. 1214-1236.
- [35] J. VON NEUMANN, *Functional operators: Measures and integrals*, vol. 1, Princeton University Press, 1950.
- [36] H. WHITNEY, *Elementary structure of real algebraic varieties*, The Annals of Mathematics, 66 (1957), pp. 545-556.
- [37] B. WIDROW, *Adaptive inverse control*, in Adaptive Systems in Control and Signal Processing 1986, Elsevier, 1987, pp. 1-5.
- [38] Y. XU, L. YU, H. XU, H. ZHANG, AND T. NGUYEN, *Vector sparse representation of color image using quaternion matrix analysis*, IEEE Transactions on image processing, 24 (2015), pp. 1315-1329.
- [39] D. YOULA, *Generalized image restoration by the method of alternating orthogonal projections*, IEEE Transactions on Circuits and Systems, 25 (1978), pp. 694-702.
- [40] L. YU, Y. XU, H. XU, AND H. ZHANG, *Quaternion-based sparse representation of color image*, in 2013 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2013, pp. 1-7.
- [41] R. ZENG, J. WU, Z. SHAO, Y. CHEN, B. CHEN, L. SENHADJI, AND H. SHU, *Color image classification via quaternion principal component analysis network*, Neurocomputing, 216 (2016), pp. 416-428.
- [42] F. ZHANG, *Quaternions and matrices of quaternions*, Linear algebra and its applications, 251 (1997), pp. 21-57.

Appendix

Proof of Lemma 2.3: For a rank- r quaternion matrix $\mathbf{E} = E_0 + E_1\mathbf{i} + E_2\mathbf{j} + E_3\mathbf{k} \in \mathcal{Q}$, by the elementary transformations it can be expressed as

$$\mathbf{E} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

where $\mathbf{A} \in \mathbb{H}^{r \times r}$ is invertible, $\mathbf{B} \in \mathbb{H}^{r \times (m-r)}$, $\mathbf{C} \in \mathbb{H}^{(n-r) \times r}$ and $\mathbf{D} \in \mathbb{H}^{(m-r) \times (n-r)}$. It is easy to find an invertible quaternion matrix

$$\mathbf{P} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B} \\ 0 & \mathbf{I}_{n-k} \end{pmatrix}$$

such that

$$\mathbf{EP} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B} \\ 0 & \mathbf{I}_{n-k} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_r & 0 \\ \mathbf{CA}^{-1} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{pmatrix},$$

with $\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} = 0$. Let

$$\mathbf{U} = \left\{ \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \in \mathbb{H}^{m \times n} : \mathbf{A} \text{ is invertible} \right\}$$

be an open set of $\mathbb{H}^{m \times n}$ which contains \mathbf{E} . Denote $\hat{\mathbb{R}}^{4m \times 4n}$ as the set of matrices that possess the structure as the real expression of an $m \times n$ quaternion matrix. By the projection ϕ defined in (3), \mathbf{U} is isomorphic to

$$\hat{U} = \left\{ \begin{pmatrix} \hat{\mathbf{A}} & \hat{\mathbf{B}} \\ \hat{\mathbf{C}} & \hat{\mathbf{D}} \end{pmatrix} \in \hat{\mathbb{R}}^{4m \times 4n} : \hat{\mathbf{A}} \in \hat{\mathbb{R}}^{4r \times 4r} \text{ is invertible} \right\},$$

which is a open subset of $\hat{\mathbb{R}}^{4m \times 4n}$. Hence, we can define $F \circ \phi : \mathbf{U} \rightarrow \hat{\mathbb{R}}^{4(m-r) \times 4(n-r)}$ as

$$F \circ \phi \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \hat{\mathbf{D}} - \hat{\mathbf{C}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}.$$

Clearly, $F \circ \phi$ is smooth. In order to show it is a submersion, we need to show $D(F \circ \phi)(\mathbf{E})$ is surjective for each $\mathbf{E} \in \mathbf{U}$. Note that $\hat{\mathbb{R}}^{4(m-r) \times 4(n-r)}$ is a vector space, the tangent vectors at $F \circ \phi(\mathbf{E})$ can be identified by the matrices in $\hat{\mathbb{R}}^{4(m-r) \times 4(n-r)}$. Given $\mathbf{E} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ and any matrix $X \in \hat{\mathbb{R}}^{4(m-r) \times 4(n-r)}$, define a curve $\tau : (-\xi, \xi) \rightarrow \hat{U}$ by

$$\tau(t) = \begin{pmatrix} \hat{\mathbf{A}} & \hat{\mathbf{B}} \\ \hat{\mathbf{C}} & \hat{\mathbf{D}} + tX \end{pmatrix}.$$

Then

$$(F \circ \phi)_* \tau'(0) = (F \circ \phi \circ \tau)'(t) = \frac{d}{dt} \Big|_{t=0} (\hat{\mathbf{D}} + tX - \hat{\mathbf{C}}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}) = X,$$

where $(F \circ \phi)_*$ is the push-forward projection relate $F \circ \phi$. Then $F \circ \phi$ is a submersion and so $\mathcal{Q} \cap \mathbf{U}$ is an embedded submanifold of $\mathbb{H}^{m \times n}$. Next, if \mathbf{E}' is an arbitrary quaternion

matrix with $\text{rank}(\mathbf{E}') = \mathbf{r}$, then it can be transformed to a quaternion matrix in \mathbf{U} by a rearrangement along its rows and columns. Let ω denote such a rearrangement which preserves the quaternion matrix rank. It follows that $\mathbf{U}_0 = \omega^{-1}(\mathbf{U})$ is a neighborhood of \mathbf{E}' and $F \circ \phi \circ \omega : \mathbf{U}_0 \rightarrow \hat{\mathbb{R}}^{4(\mathbf{m}-\mathbf{r}) \times 4(\mathbf{n}-\mathbf{r})}$ is a submersion whose zero level set is $\mathcal{Q} \cap \mathbf{U}_0$. Thus every point in \mathcal{Q} has a neighborhood $\mathbf{U}_0 \subseteq \mathbb{H}^{m \times n}$ such that $\mathcal{Q} \cap \mathbf{U}_0$ is an embedded submanifold of \mathbf{U}_0 , so \mathcal{Q} is an embedded submanifold. Moreover, note that $\dim((F \circ \phi)_* \tau'(0)) = 4(m+n)r - 4r^2$ which is saying that \mathcal{Q} possess dimension $4(m+n)r - 4r^2$. \square

Proof of Proposition 5.1: Denote $\mathbf{A}^{*ij} = A_0 - A_1\mathbf{i} - A_2\mathbf{j} + A_3\mathbf{k}$ then for two arbitrary quaternion matrices \mathbf{B}, \mathbf{C} , we have $(\mathbf{BC})^{*ij} = \mathbf{B}^{*ij}\mathbf{C}^{*ij}$. Similarly, we have $(\mathbf{BC})^{*ik} = \mathbf{B}^{*ik}\mathbf{C}^{*ik}$ and $(\mathbf{BC})^{*jk} = \mathbf{B}^{*jk}\mathbf{C}^{*jk}$. Note that $\text{rank}(\mathbf{A}) = r$, then there exist a full column rank matrix $\mathbf{U} \in \mathbb{H}^{m \times r}$ and a full row rank matrix $\mathbf{V} \in \mathbb{H}^{r \times n}$ such that $\mathbf{A} = \mathbf{U} \cdot \mathbf{V}$. Moreover,

$$\mathbf{A}^{*ij} = \mathbf{U}^{*ij} \cdot \mathbf{V}^{*ij}, \mathbf{A}^{*ik} = \mathbf{U}^{*ik} \cdot \mathbf{V}^{*ik}, \text{ and } \mathbf{A}^{*jk} = \mathbf{U}^{*jk} \cdot \mathbf{V}^{*jk}.$$

For the pure quaternion part of \mathbf{A} , we have

$$\begin{aligned} \mathbf{A}_p &= \frac{1}{4}((\mathbf{A} - \mathbf{A}^{*ij}) + (\mathbf{A} - \mathbf{A}^{*jk}) + (\mathbf{A} - \mathbf{A}^{*ik})) = \frac{1}{4}(3\mathbf{A} - \mathbf{A}^{*ij} - \mathbf{A}^{*ik} - \mathbf{A}^{*jk}) \\ &= \frac{1}{4}(3\mathbf{UV} - \mathbf{U}^{*ij}\mathbf{V}^{*ij} - \mathbf{U}^{*ik}\mathbf{V}^{*ik} - \mathbf{U}^{*jk}\mathbf{V}^{*jk}) \\ &= \frac{1}{4}(3\mathbf{U}, \mathbf{U}^{*ij}, \mathbf{U}^{*ik}, \mathbf{U}^{*jk}) \begin{pmatrix} \mathbf{V} \\ -\mathbf{V}^{*ij} \\ -\mathbf{V}^{*ik} \\ -\mathbf{V}^{*jk} \end{pmatrix}. \end{aligned}$$

Then $r \leq \text{rank}(\mathbf{A}_p) \leq 4r$. This completes the proof. \square