Quaternion Matrix Optimization and The Underlying Calculus

Liqun Qi\textsuperscript{*} Ziyuan Luo\textsuperscript{†} Qing-Wen Wang\textsuperscript{‡} and Xinzhen Zhang\textsuperscript{§}

October 29, 2020

Abstract

Optimization models involving quaternion matrices are widely used in color image process and other engineering areas. These models optimize real functions of quaternion matrix variables. In particular, 0-norms and rank functions of quaternion matrices are not continuous at all. Yet calculus with derivatives, subdifferentials and generalized subdifferentials of such real functions is needed to handle such models. In this paper, we introduce first and second order derivatives and establish their calculation rules for such real functions. Our approach is consistent with the subgradient concept for norms of quaternion matrix variables, recently introduced in the literature. We develop the concepts of generalized subdifferentials of lower semi-continuous functions of quaternion matrices, and use them to analyze the optimality conditions of a sparse low rank color image denoising model. We introduce R-product for two quaternion matrix vectors, as a key tool for our calculus. We show that the real representation set of low-rank quaternion matrices is closed and semi-algebraic. We also establish first order and second order optimality conditions for constrained optimization problems of real functions in quaternion matrix variables.

Key words. real functions, quaternion matrix variables, gradients, subdifferentials, generalized subdifferentials, color image inpainting, color image denoising.

\textsuperscript{*}Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maqilq@polyu.edu.hk).
\textsuperscript{†}Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China. (zyluo@bjtu.edu.cn). This author’s work was supported by NSFC (Grant No. 11771038) and Beijing Natural Science Foundation (Grant No. Z190002).
\textsuperscript{‡}Department of Mathematics, Shanghai University, Shanghai 200444, China. (wqw@t.shu.edu.cn). This author’s work was supported by NSFC (Grant No. 11971294).
\textsuperscript{§}Corresponding author. School of Mathematics, Tianjin University, Tianjin 300354 China; (xzzhang@tju.edu.cn). This author’s work was supported by NSFC (Grant No. 11871369).
# 1 Introduction

Quaternion matrix methods have been widely used in color image processing, including color image denoising and inpainting, color face recognition, etc., \[4, 8, 10, 15, 16, 17, 19, 22\]. Quaternions have also been widely used in the other engineering areas \[18\]. Optimization techniques are frequently used in this process.

Various real functions of quaternion matrix variables, such as Frobenius norms, nuclear norms, spectral norms, 1-norms, traces, 0-norms and rank functions of quaternion matrices, arise from engineering applications.

To optimize such real functions, derivatives, subdifferentials and generalized subdifferentials are needed to handle them.

For the color image denoising problem, optimization models may involve 0-norms and rank functions of quaternion matrix variables. They are not continuous at all. In sparse optimization \[2, 11\], generalized subdifferentials of 0-norms and rank functions of real matrix variables are used. This uses the knowledge of variational analysis \[5, 7, 9, 13, 14\]. In this paper, we aim to develop adequate analysis tools for developing derivatives, subdifferentials and generalized subdifferentials of real functions of quaternion matrix variables, such that we can establish optimality conditions of such optimization models and prepare bases for future works on convergence analysis of algorithms for such optimization models.

In the literature, there are some works on derivatives of real functions of quaternion variables \[12, 18\], and subgradients of norms of quaternion matrix variables \[8\]. There are no discussion on generalized subdifferentials of 0-norms and rank functions of quaternion matrix variables.

In \[3\], Chen, Qi, Zhang and Xu formulated the color image inpainting problem as an equality constrained optimization problem of real functions in quaternion matrix variables, and proposed a lower rank quaternion decomposition (LRQD) algorithm to solve it. To conduct convergence analysis for their algorithm, they introduced a concise form

\[
\nabla f(X) = \frac{\partial f}{\partial X_0} + i \frac{\partial f}{\partial X_1} + j \frac{\partial f}{\partial X_2} + k \frac{\partial f}{\partial X_3}
\]

for the gradient of a real function \(f\) in a quaternion matrix variable \(X\). This form is different from the generalized HR calculus studied in \[18\]. The first order optimality necessary condition of their quaternion least squares problem has a simple expression with this form. With this tool, convergence and convergence rate of their algorithm are established.

This motivates us to study constrained optimization of real functions in quaternion matrix variables systematically. We consider general inequality and equality constrained optimization problems of real functions in quaternion matrix variables. We use the LRQD problem as a prototype problem. We introduce R-product and R-linear independence of quaternion matrix vectors, and establish the first order necessary optimality condition for the general inequality and equality constrained optimization problem of real functions in quaternion matrix variables. We present a method to calculate the second order partial derivatives of real functions in
quaternion matrix variables, and establish the second order necessary optimality condition and second order sufficient optimality condition for the general constrained optimization problem of real functions in quaternion matrix variables.

Norms of quaternion matrix variables may not be continuously differentiable, but they are always convex. In 2019, Jia, Ng and Song [8] introduced subgradients for norms of quaternion matrix variables.

In this paper, we show that our approach is consistent with the subgradient concept for norms of quaternion matrix variables, introduced in [8]. We also establish the relations between subgradients and derivatives of real functions of quaternion matrix variables.

The 0-norms and the ranks of quaternion matrices are not continuous at all, but they are lower semi-continuous. They are very useful in applications [4, 19]. In this paper, we develop the concepts of generalized subdifferentials of lower semi-continuous functions of quaternion matrices, and use them to analyze the optimality conditions of a sparse low rank color image denoising model. We show that the real representation set of low-rank quaternion matrices is closed and semi-algebraic.

The generalized subdifferential calculus is totally new in the literature and useful in color image applications.

In the next section, we present some necessary preliminary knowledge on quaternions and quaternion matrices. The general inequality and equality constrained optimization problem of real functions in quaternion matrix variables and its LRQD prototype problem are presented in Section 3. In Section 4, we introduce R-product, R-linear independence and present first order necessary optimality condition for the general constrained optimization problem. We also establish the product rule and the chain rule of first order derivatives. We study second order derivatives and present second order optimality conditions in Section 5. In Section 6, we study convex functions of quaternion matrix variables, their subdifferentials, and the relations with our derivatives of real functions of quaternion matrix variables. We introduce the generalized subdifferentials of lower semi-continuous functions of quaternion matrices, and use them to analyze the optimality conditions of a sparse low rank color image denoising model in Section 7. Some final remarks are made in Section 8.

2 Preliminary

2.1 Quaternions

In general, we use the notation in [21, 20]. We denote the real field, the complex field and the quaternion algebra by \( \mathbb{R} \), \( \mathbb{C} \) and \( \mathbb{Q} \), respectively. Scalars, vectors, matrices and tensors are denoted by small letters, bold small letters, capital letters and calligraphic letters, respectively. Vectors with matrix components are denoted by bold capital letters. For example, we have \( \mathbf{X} = (W, Y, Z) \). They are called matrix component vectors. We use \( \mathbf{0}, O, \mathcal{O} \) and \( \mathbf{O} \) to denote zero
vector, zero matrix, zero tensor and zero matrix component vector with adequate dimensions. The three imaginary units of quaternions are denoted by $i, j$ and $k$. We have
\[ i^2 = j^2 = k^2 = ijk = -1, \]
\[ ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j. \]
These rules, along with the distribution law, determine the product of two quaternions. The multiplication of quaternions is noncommutative.

Let $x = x_0 + x_1i + x_2j + x_3k \in \mathbb{Q}$, where $x_0, x_1, x_2, x_3 \in \mathbb{R}$. The conjugate of $x$ is
\[ x^* = x_0 - x_1i - x_2j - x_3k, \]
the modulus of $x$ is
\[ |x| = |x^*| = \sqrt{xx^*} = \sqrt{x^*x} = \sqrt{x_0^2 + x_1^2 + x_2^2 + x_3^2}, \]
and if $x \neq 0$, then $x^{-1} = \frac{x^*}{|x|^2}$.

### 2.2 Quaternion Matrices

The collections of real, complex and quaternion $m \times n$ matrices are denoted by $\mathbb{R}^{m \times n}$, $\mathbb{C}^{m \times n}$ and $\mathbb{Q}^{m \times n}$, respectively.

A quaternion matrix $A = (a_{ij}) \in \mathbb{Q}^{m \times n}$ can be denoted as
\[ A = A_0 + A_1i + A_2j + A_3k, \tag{1} \]
where $A_0, A_1, A_2, A_3 \in \mathbb{R}^{m \times n}$. The transpose of $A$ is $A^\top = (a_{ji})$. The conjugate of $A$ is $\bar{A} = (a_{ij}^*)$. The conjugate transpose of $A$ is $A^* = (a_{ji}^*) = \bar{A}^\top$. For $A, B \in \mathbb{Q}^{m \times n}$, their inner product is defined as
\[ \langle A, B \rangle = \text{Tr}(A^*B), \]
where $\text{Tr}(A^*B)$ denotes the trace of $A^*B$. The Frobenius norm of $A$ is
\[ \|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\text{Tr}(A^*A)} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}. \]

The following theorem for the QSVD of a quaternion matrix was proved by Zhang [21].

**Theorem 2.1. (Zhang 1997)** Any quaternion matrix $A \in \mathbb{Q}^{m \times n}$ has the following QSVD form
\[ X = U \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V^*, \tag{2} \]
where $U \in \mathbb{Q}^{m \times m}$ and $V \in \mathbb{Q}^{n \times n}$ are unitary, and $\Sigma_r = \text{diag}\{\sigma_1, \cdots, \sigma_r\}$ is a real nonnegative $r \times r$ diagonal matrix, with $\sigma_1 \geq \cdots \geq \sigma_r$ as the singular values of $A$.
The 1-norm of $A = (a_{ij}) \in \mathbb{Q}^{m \times n}$ is defined by $\|A\|_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|$. The $\infty$-norm of $A$ is defined by $\|A\|_\infty = \max_{i,j} |a_{ij}|$. The spectral norm of $A$ is defined as $\|A\|_2 = \max \{\sigma_1, \ldots, \sigma_r\}$. The nuclear norm of $A$ is defined as $\|A\|_* = \sum_{i=1}^{r} \sigma_i$.

The 0-norm of $A = (a_{ij}) \in \mathbb{Q}^{m \times n}$ is defined by $\|A\|_0 = \text{the number of } a_{ij} \neq 0$. It is not a true norm as it does not satisfy the triangular inequality of norms, and it is not continuous at all. However, it plays an important role in sparse color image processing \cite{19}. In some papers, it is called the counting function \cite{9}.

The rank of $A \in \mathbb{Q}^{m \times n}$, denoted as rank($A$), is equal to the number of nonzero singular values of $A$. It is also not continuous. Yet it plays an important role in color image inpainting \cite{4}. In \cite{7}, the rank function of real matrices was shown to be lower semi-continuous. With a similar argument, it is seen that the rank function of quaternion matrices are also semicontinuous.

However, the 0-norm and the rank of $A$ is lower semi-continuous with respect to $A$. In Section 7, we will analyze their generalized subdifferentials.

For a quaternion matrix $A = A_0 + A_1i + A_2j + A_3k \in \mathbb{Q}^{m \times n}$, its real representation \cite{20} is

$$A^R = \begin{pmatrix} A_0 & -A_1 & -A_2 & -A_3 \\ A_1 & A_0 & -A_3 & A_2 \\ A_2 & A_3 & A_0 & -A_1 \\ A_3 & -A_2 & A_1 & A_0 \end{pmatrix}.$$  

A color image dataset can be represented as a pure quaternion matrix

$$A = A_1i + A_2j + A_3k \in \mathbb{Q}^{m \times n},$$

where $A_1, A_2$ and $A_3$ are the three real $m \times n$ matrices.

We use $O_{m \times n}$ to denote the zero matrix in $\mathbb{Q}^{m \times n}$.

More knowledge of quaternion matrices can be found in \cite{20, 21}.

3 The General Problem and The Prototype Problem

We consider matrix component vector $X \equiv (W, Y, Z)$, where $W \in \mathbb{Q}^{m_1 \times n_1}$, $Y \in \mathbb{Q}^{m_2 \times n_2}$ and $Z \in \mathbb{Q}^{m_3 \times n_3}$. We may consider matrix component vectors with more components. But three components are enough for illustrating the problem. The general constrained optimization problem of quaternion matrix variables has the following form

$$\min \left \{ f(X) : h_j(X) = 0, j = 1, \cdots, p, g_k(X) \leq 0, k = 1, \cdots, q \right \},$$

where $f, h_j, g_k : \mathbb{Q}^{m_1 \times n_1} \times \mathbb{Q}^{m_2 \times n_2} \times \mathbb{Q}^{m_3 \times n_3} \to \mathbb{R}$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$.

Denote $\mathbb{H} := \mathbb{Q}^{m_1 \times n_1} \times \mathbb{Q}^{m_2 \times n_2} \times \mathbb{Q}^{m_3 \times n_3}$.

Suppose that

$$W = W_0 + W_1i + W_2j + W_3k,$$

$$Y = Y_0 + Y_1i + Y_2j + Y_3k,$$

$$Z = Z_0 + Z_1i + Z_2j + Z_3k,$$
where \( W_i \in \mathbb{R}^{m_i \times n_1} \), \( Y_i \in \mathbb{R}^{m_2 \times n_2} \) and \( Z_i \in \mathbb{R}^{m_3 \times n_3} \), for \( i = 0, 1, 2, 3 \). Then \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) can be regarded as functions of \( W_i, Y_i \) and \( Z_i \) for \( i = 0, 1, 2, 3 \), and denote such functions as \( f^R, h_j^R \) and \( g_k^R \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \). We call them the real representations of \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \).

For \( \mathbf{X} = (W, Y, Z) \), denote \( R(\mathbf{X}) = (W_0, W_2, W_3, Y_0, Y_1, Y_2, Y_3, Z_0, Z_1, Z_2, Z_3) \), and call \( R(\mathbf{X}) \) the real representation of \( \mathbf{X} \).

We have the following three different assumptions.

**Normal Assumption** The functions \( f^R, h_j^R \) and \( g_k^R \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are locally Lipschitz continuous with respect to \( W_i, Y_i \) and \( Z_i \) for \( i = 0, 1, 2, 3 \).

In this case, we say that \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are locally Lipschitz. All the norms of quaternion matrices satisfy this assumption.

**Middle Assumption** The functions \( f^R, h_j^R \) and \( g_k^R \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are continuously differentiable with respect to \( W_i, Y_i \) and \( Z_i \) for \( i = 0, 1, 2, 3 \).

In this case, we say that \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are continuously differentiable.

**Strong Assumption** The functions \( f^R, h_j^R \) and \( g_k^R \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are twice continuously differentiable with respect to \( W_i, Y_i \) and \( Z_i \) for \( i = 0, 1, 2, 3 \).

In this case, we say that \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) are twice continuously differentiable. The squares of the Frobenius norms of quaternion matrices satisfy this assumption.

With the middle or strong assumption, the general constrained optimization problem (3) can be solved by standard optimization methods. This is somewhat tedious. In this paper, we attempt to find some concise quaternion forms for studying (3).

We also have a prototype constrained optimization problem

\[
\min \left\{ \| YZ - W \|_F^2 : W_\Omega = D_\Omega \right\},
\]

where \( W, D \in \mathbb{Q}^{m \times n} \), \( Y \in \mathbb{Q}^{m \times r} \) and \( Z \in \mathbb{Q}^{r \times n} \), \( D \) is a quaternion dataset matrix, \( \Omega \) denotes the set of \( m \times n \) quaternion matrices, \( W \) is the observed matrix, \( r \) is the upper bound of the rank of the low rank decomposition \( YZ \) to approximate \( W \).

This constrained optimization problem arises from the color image inpainting problem (3).

Comparing with (3), for (4), we have \( m_1 = m_2 = m \), \( n_1 = n_3 = n \), \( n_2 = m_3 = r \), \( p = |\Omega| \), the cardinality of \( \Omega \), and \( q = 0 \). Clearly, (4) satisfies the strong assumption.

We will use (4) as an example for the optimality conditions and algorithms we study for (3).
4 R-Product, R-Linear Independence and First Order Derivative

Suppose that $A$ and $E$ are two real matrices with the same dimension. For example, assume that $A = (a_{ij})$ and $E = (e_{ij})$ are $m \times n$ real matrices. Define the R-product (real product) of $A$ and $E$ as

$$ A \cdot E := \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} e_{ij}. $$

Consider the matrix component vector space $H \equiv Q^{m_1 \times n_1} \times Q^{m_2 \times n_2} \times Q^{m_3 \times n_3}$.

Suppose that $A \equiv (B, C, D) \in H$ and $H \equiv (E, F, G) \in H$. Assume that $B = B_0 + B_1 i + B_2 j + B_3 k, \ldots, G = G_0 + G_1 i + G_2 j + G_3 k$, where $B_i, C_i, D_i, E_i, F_i$ and $G_i$ are real matrices with corresponding dimensions. Define the R-product (real product) of $A$ and $H$ as

$$ A \cdot H := \sum_{i=0}^{3} (B_i \cdot E_i + C_i \cdot F_i + D_i \cdot G_i). $$

For $A \equiv (B, C, D) \in H$ and $H \equiv (E, F, G) \in H$, we may define the inner product of $A$ and $H$ as

$$ \langle A, H \rangle = \langle B, E \rangle + \langle C, F \rangle + \langle D, G \rangle. $$

We have the following proposition.

Proposition 4.1. Suppose that $A \equiv (B, C, D) \in H$ and $H \equiv (E, F, G) \in H$. Then the R-product of $A$ and $H$ is the real part of $\langle A, H \rangle$.

Proof. Let

$$ B = B_0 + B_1 i + B_2 j + B_3 k, \quad E = E_0 + E_1 i + E_2 j + E_3 k, $$

$$ B_i = (b_{pqi}), \quad E_i = (e_{pqj}), \quad \text{for} \quad i = 0, 1, 2, 3. $$

Then

$$ \langle B, E \rangle = \text{Tr}(B^* E) = \sum_{p=1}^{m_1} \sum_{q=1}^{n_1} (b_{pq0} - b_{pq1} i - b_{pq2} j - b_{pq3} k) (e_{pq0} + e_{pq1} i + e_{pq2} j + e_{pq3} k). $$

We see that the real part of $(b_{pq0} - b_{pq1} i - b_{pq2} j - b_{pq3} k) (e_{pq0} + e_{pq1} i + e_{pq2} j + e_{pq3} k)$,

$$ \text{Re} \left( (b_{pq0} - b_{pq1} i - b_{pq2} j - b_{pq3} k) (e_{pq0} + e_{pq1} i + e_{pq2} j + e_{pq3} k) \right) = \sum_{i=0}^{3} b_{pq0} e_{pq1}. $$

This implies that the real part of $\langle B, E \rangle$ is $\sum_{i=0}^{3} B_i \cdot E_i$. Hence, the real part of $\langle A, H \rangle$ is $A \cdot H$. \qed
This proposition reveals the meaning of the R-product. In Section 7, it connects our derivative concept with the subgradient concept of Jia, Ng and Song [8].

Suppose that \( A^{(j)} \equiv (A^{(j)}, B^{(j)}, C^{(j)}) \in \mathbb{H} \) for \( j = 1, \cdots, p \). We say that \( \{ A^{(j)} : j = 1, \cdots, p \} \) is R-linearly dependent if there are real numbers \( \alpha_j \) for \( j = 1, \cdots, p \) such that some of them are nonzero and

\[
\sum_{j=1}^{p} \alpha_j A^{(j)} = 0.
\]

Suppose that \( f(X) : \mathbb{H} \to \mathbb{R} \) satisfies the normal assumption, with \( X = (W, Y, Z) \), and \( f \) is differentiable at \( X \) with respect to \( W_i, Y_i \) and \( Z_i \) for \( i = 0, 1, 2, 3 \). Then we define the partial derivatives and gradient of \( f \) at \( X \) as:

\[
\frac{\partial}{\partial W} f(X) = \frac{\partial}{\partial W_0} f(X) + \frac{\partial}{\partial W_1} f(X) i + \frac{\partial}{\partial W_2} f(X) j + \frac{\partial}{\partial W_3} f(X) k,
\]

\[
\frac{\partial}{\partial Y} f(X) = \frac{\partial}{\partial Y_0} f(X) + \frac{\partial}{\partial Y_1} f(X) i + \frac{\partial}{\partial Y_2} f(X) j + \frac{\partial}{\partial Y_3} f(X) k,
\]

\[
\frac{\partial}{\partial Z} f(X) = \frac{\partial}{\partial Z_0} f(X) + \frac{\partial}{\partial Z_1} f(X) i + \frac{\partial}{\partial Z_2} f(X) j + \frac{\partial}{\partial Z_3} f(X) k,
\]

\[
\nabla f(X) = \left( \frac{\partial}{\partial W} f(X), \frac{\partial}{\partial Y} f(X), \frac{\partial}{\partial Z} f(X) \right).
\]

We see that \( \nabla f(X) \in \mathbb{H} \).

Define the directional derivative of \( f \) at \( X = (W, Y, Z) \in \mathbb{H} \) in the direction \( \Delta X = (\Delta W, \Delta Y, \Delta Z) \in \mathbb{H} \) as

\[
f'(X; \Delta X) = \lim_{t \to 0} \frac{f(X + t \Delta X) - f(X)}{t}.
\]

Note that while the gradient of \( f \) is in \( \mathbb{H} \), the directional derivative of \( f \) is real. They are connected via the R-product operation in \( \mathbb{H} \).

**Proposition 4.2.** Suppose that \( X = (W, Y, Z) \in \mathbb{H} \) and \( \Delta X = (\Delta W, \Delta Y, \Delta Z) \in \mathbb{H} \). Then

\[
f'(X; \Delta X) = \nabla f(X) \cdot \Delta X.
\]

Furthermore, if \( f \) satisfies the middle assumption, then we have

\[
f(X + \Delta X) = f(X) + \nabla f(X) \cdot \Delta X + o(\|\Delta X\|_F).
\]

**Proof.** We may regard \( f(X) \) as a function of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \). Then

\[
f'(X; \Delta X) = \sum_{i=0}^3 \left[ \frac{\partial}{\partial W_i} f(X) \Delta W_i + \frac{\partial}{\partial Y_i} f(X) \Delta Y_i + \frac{\partial}{\partial Z_i} f(X) \Delta Z_i \right] = \nabla f(X) \cdot \Delta X.
\]
Furthermore,

\[ f(X + \Delta X) - f(X) = \sum_{i=0}^{3} \left[ \frac{\partial}{\partial W_i} f(X) \Delta W_i + o(\|\Delta W_i\|) + \frac{\partial}{\partial Y_i} f(X) \Delta Y_i + o(\|\Delta Y_i\|) + \frac{\partial}{\partial Z_i} f(X) \Delta Z_i + o(\|\Delta Z_i\|) \right] \]

\[ = \nabla f(X) : \Delta X + o(\|\Delta X\|). \]

Now, we may study the first order optimality conditions of (3).

**Theorem 4.3.** Suppose that the functions \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) satisfy the middle assumption. Assume that \( X^# = (W^#, Y^#, Z^#) \in \mathbb{H} \) is an optimal solution of (3). If

\[ \{ \nabla h_j (X^#), j = 1, \cdots, p \} \bigcup \{ \nabla g_k (X^#) : g_k(X^#) = 0, 1 \leq k \leq q \} \]  \hspace{1cm} (5)

is \( R \)-linearly independent, then there are Lagrangian multipliers \( \lambda_j, \mu_k \in \mathbb{R} \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \), such that

\[ \nabla f (X^#) + \sum_{j=1}^{p} \lambda_j \nabla h_j (X^#) + \sum_{k=1}^{p} \mu_k \nabla g_k (X^#) = 0, \]  \hspace{1cm} (6)

\[ h_j (X^#) = 0, \quad j = 1, \cdots, p, \]  \hspace{1cm} (7)

\[ g_k (X^#) \leq 0, \quad \mu_k \geq 0, \quad \mu_k g_k (X^#) = 0, \quad k = 1, \cdots, q. \]  \hspace{1cm} (8)

**Proof.** Again, we may regard \( f, h_j, g_k \) for \( j = 1, \cdots, p \), \( k = 1, \cdots, q \) as functions of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \). Then (3) is equivalent to linear independence constraint qualification of such an optimization problem of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \). Then from the first order optimality condition for real constrained optimization, we have (6-8). \( \square \)

Note that the Lagrangian multipliers are real numbers. We call \( X^# \in \mathbb{H} \), which satisfies (6-8) with some Lagrangian multipliers, a stationary point of (3). We may reduce the \( R \)-linearly independence condition in Theorem 4.3 to other constraint qualifications for nonlinear programs.

We now consider the prototype problem (1). Let \( f(X) \equiv \frac{1}{2}\|YZ - W\|^2_F \). Then by (3), we have

\[ \frac{\partial}{\partial W} f(X) = W - YZ, \]  \hspace{1cm} (9)

\[ \frac{\partial}{\partial Y} f(X) = (YZ - W)Z^*, \]  \hspace{1cm} (10)

\[ \frac{\partial}{\partial Z} f(X) = Y^*(YZ - W) \]  \hspace{1cm} (11)

and the following theorem.
Theorem 4.4. Assume that $X^# = (W^#, Y^#, Z^#) \in \mathbb{H}$ is an optimal solution of $[4]$. Then $X^#$ is a stationary point of $[4]$, i.e.,

$$W_{\Omega_c}^# = (A^#B^#)_{\Omega_c},$$

$$(A^#B^# - W^#)(B^#)^* = O_{mxr},$$

$$(A^#)^*(A^#B^# - W^#) = O_{rnx},$$

$$W_{\Omega^c}^# = D_{\Omega},$$

where $\Omega_c$ is the complement set of $\Omega$.

We now study the product rule and the chain rule of first order derivatives.

Theorem 4.5. Suppose that $f(X), g(X) : \mathbb{H} \to \mathbb{R}$ satisfies the middle assumption, with $X = (W, Y, Z)$. Then

$$\nabla(f(X)g(X)) = f(X)\nabla g(X) + g(X)\nabla f(X).$$

(12)

Proof. We have

$$\frac{\partial}{\partial W}(f(X)g(X))
= \frac{\partial}{\partial W_0}(f(X)g(X)) + \frac{\partial}{\partial W_1}(f(X)g(X))i + \frac{\partial}{\partial W_2}(f(X)g(X))j + \frac{\partial}{\partial W_3}(f(X)g(X))k
= \left[f(X)\frac{\partial}{\partial W_0}g(X) + g(X)\frac{\partial}{\partial W_0}f(X)\right] + \left[f(X)\frac{\partial}{\partial W_1}g(X) + g(X)\frac{\partial}{\partial W_1}f(X)\right]i
+ \left[f(X)\frac{\partial}{\partial W_2}g(X) + g(X)\frac{\partial}{\partial W_2}f(X)\right]j + \left[f(X)\frac{\partial}{\partial W_3}g(X) + g(X)\frac{\partial}{\partial W_3}f(X)\right]k
= f(X)\frac{\partial}{\partial W}g(X) + g(X)\frac{\partial}{\partial W}f(X).$$

Similarly, we have

$$\frac{\partial}{\partial Y}(f(X)g(X)) = f(X)\frac{\partial}{\partial Y}g(X) + g(X)\frac{\partial}{\partial Y}f(X)$$

and

$$\frac{\partial}{\partial Z}(f(X)g(X)) = f(X)\frac{\partial}{\partial Z}g(X) + g(X)\frac{\partial}{\partial Z}f(X).$$

Then

$$\nabla(f(X)g(X))
= \left(\frac{\partial}{\partial W}(f(X)g(X)), \frac{\partial}{\partial Y}(f(X)g(X)), \frac{\partial}{\partial Z}(f(X)g(X))\right)
= \left(f(X)\frac{\partial}{\partial W}g(X) + g(X)\frac{\partial}{\partial W}f(X), f(X)\frac{\partial}{\partial Y}g(X) + g(X)\frac{\partial}{\partial Y}f(X), f(X)\frac{\partial}{\partial Z}g(X) + g(X)\frac{\partial}{\partial Z}f(X)\right)
= f(X)\nabla g(X) + g(X)\nabla f(X).$$

\[\square\]

Similarly, we may prove the following theorem.
Theorem 4.6. Suppose that $f(X) : \mathbb{H} \rightarrow \mathbb{R}$ satisfies the middle assumption, with $X = (W, Y, Z)$, and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is continuously differentiable. Then
\[
\nabla \phi(f(X)) = \phi'(f(X)) \nabla f(X).
\] (13)

In [3], the Kurdyka-Lojasiewicz inequality was established for real functions of quaternion matrix variables.

5 Second Order Derivative

Suppose that $f(X) \equiv f(W, Y, Z) : \mathbb{H} \rightarrow \mathbb{R}$ is twice continuously differentiable in the sense of the strong assumption. Then the second order derivative of $f$ exists. We first consider the second order partial derivatives of $f$. To make an example, we consider $\partial^2 \partial Y \partial W f(X) \Delta Y$. It is more convenient to consider $\partial^2 \partial Y \partial W f(X) \Delta Y$ and $\partial^2 \partial Y \partial W f(X) \Delta Y \cdot \Delta W$.

Let $\frac{\partial}{\partial W} f(X)$ be as considered in the last section. Let $\Delta Y \in \mathbb{Q}^{m_2 \times n_2}$. Suppose that
\[
\frac{\partial}{\partial W} f(W, Y + \Delta Y, Z) - \frac{\partial}{\partial W} f(W, Y, Z) = \phi(W, Y + \Delta Y, Z) + \eta(W, Y + \Delta Y, Z),
\]
where $\phi$ is R-linear in $\Delta Y$ in the sense that for any $\alpha, \beta \in \mathbb{R}$ and $\Delta Y^{(1)}, \Delta Y^{(2)} \in \mathbb{Q}^{m_2 \times n_2}$,
\[
\phi \left(W, Y + \alpha \Delta Y^{(1)} + \beta \Delta Y^{(2)}, Z\right) = \alpha \phi \left(W, Y + \alpha \Delta Y^{(1)}, Z\right) + \beta \phi \left(W, Y + \alpha \Delta Y^{(2)}, Z\right),
\]
and
\[
\eta(W, Y + \Delta Y, Z) = o \left(\|\Delta Y\|_F\right),
\]
i.e.,
\[
\lim_{\|\Delta Y\|_F \rightarrow 0} \frac{\eta(W, Y + \Delta Y, Z)}{\|\Delta Y\|_F} = 0.
\]
Then we have
\[
\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y = \phi(W, Y + \Delta Y, Z).
\]
Later, we will see that we may use this approach to calculate $\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y \cdot \Delta W$, for $f$ defined by our prototype problem (4).

We may express the other second order partial derivatives of $f$ similarly.

Proposition 5.1. Under the strong assumption, we have
\[
\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y \cdot \Delta W = \frac{\partial^2}{\partial W \partial Y} f(X) \Delta W \cdot \Delta Y,
\]
\[
\frac{\partial^2}{\partial Z \partial W} f(X) \Delta Z \cdot \Delta W = \frac{\partial^2}{\partial W \partial Z} f(X) \Delta W \cdot \Delta Z,
\]
and
\[
\frac{\partial^2}{\partial Y \partial Z} f(X) \Delta Y \cdot \Delta Z = \frac{\partial^2}{\partial Z \partial Y} f(X) \Delta Z \cdot \Delta Y.
\]
Proof. Let $\Delta W = \Delta W_0 + \Delta W_1 + \Delta W_2 + \Delta W_3 \in \mathbb{Q}_m^{n_1}$ and $\Delta Y = \Delta Y_0 + \Delta Y_1 + \Delta Y_2 + \Delta Y_3 \in \mathbb{Q}_m^{n_2}$. Then $\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y \cdot \Delta W$ and $\frac{\partial^2}{\partial W \partial Y} f(X) \Delta W \cdot \Delta Y$ are real. We have

$$\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y \cdot \Delta W = \sum_{i,j=0}^3 \frac{\partial^2}{\partial Y_i \partial W_j} f(X) \Delta Y_i \cdot \Delta W_j$$

$$= \sum_{i,j=0}^3 \frac{\partial^2}{\partial W_j \partial Y_i} f(X) \Delta W_j \cdot \Delta Y_i = \frac{\partial^2}{\partial W \partial Y} f(X) \Delta W \cdot \Delta Y.$$ 

The other two equalities can be proved similarly. \hfill \Box

Then we may define $\nabla^2 f(X)$ by

$$\frac{1}{2} \nabla^2 f(X) \Delta X \cdot \Delta X = \frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y \cdot \Delta W + \frac{\partial^2}{\partial Z \partial W} f(X) \Delta Z \cdot \Delta W + \frac{\partial^2}{\partial Y \partial Z} f(X) \Delta Y \cdot \Delta Z$$

$$+ \frac{1}{2} \frac{\partial^2}{\partial W^2} f(X) \Delta W \cdot \Delta W + \frac{1}{2} \frac{\partial^2}{\partial Y^2} f(X) \Delta Y \cdot \Delta Y + \frac{1}{2} \frac{\partial^2}{\partial Z^2} f(X) \Delta Z \cdot \Delta Z.$$ 

Here, $\Delta X = (\Delta W, \Delta Y, \Delta Z) \in \mathbb{H}$.

If $\frac{1}{2} \nabla^2 f(X) \Delta X \cdot \Delta X \geq 0$ for any $\Delta X \in \mathbb{H}$, then we say that $\nabla^2 f$ is positive semi-definite at $X$. If $\frac{1}{2} \nabla^2 f(X) \Delta X \cdot \Delta X > 0$ for any $\Delta X \in \mathbb{H}$ and $\Delta X \neq 0$, then we say that $\nabla^2 f$ is positive definite at $X$.

**Proposition 5.2.** Under the strong assumption, we have

$$f(X + \Delta X) = f(X) + \nabla f(X) \cdot \Delta X + \frac{1}{2} \nabla^2 f(X) \Delta X \cdot \Delta X + o(\|\Delta X\|^2).$$ (14)

Proof. Again, we may regard $f(X)$ as a function of $W_i, Y_i$ and $Z_i$, $i = 0, 1, 2, 3$. Then

$$f(X + \Delta X) - f(X) = \sum_{i=0}^3 \left[ \frac{\partial}{\partial W_i} f(X) \Delta W_i + \frac{\partial}{\partial Y_i} f(X) \Delta Y_i + \frac{\partial}{\partial Z_i} f(X) \Delta Z_i \right] + \sum_{i,j=0}^3 \frac{\partial^2}{\partial Y_i \partial W_j} f(X) \Delta Y_i \cdot \Delta W_j$$

$$+ \sum_{i,j=0}^3 \frac{\partial^2}{\partial Z_i \partial W_j} f(X) \Delta Z_i \cdot \Delta W_j + \sum_{i,j=0}^3 \frac{\partial^2}{\partial Y_i \partial Z_j} f(X) \Delta Y_i \cdot \Delta Z_j + \frac{1}{2} \sum_{i,j=0}^3 \frac{\partial^2}{\partial W_i \partial W_j} f(X) \Delta W_i \cdot \Delta W_j$$

$$+ \frac{1}{2} \sum_{i,j=0}^3 \frac{\partial^2}{\partial Y_i \partial Y_j} f(X) \Delta Y_i \cdot \Delta Y_j + \frac{1}{2} \sum_{i,j=0}^3 \frac{\partial^2}{\partial Z_i \partial Z_j} f(X) \Delta Z_i \cdot \Delta Z_j$$

$$+ \sum_{i=0}^3 \left[ o(\|\Delta W_i\|^2) + o(\|\Delta Y_i\|^2) + o(\|\Delta Z_i\|^2) \right]$$

$$= \nabla f(X) \cdot \Delta X + \frac{1}{2} \nabla^2 f(X) \Delta X \cdot \Delta X + o(\|\Delta X\|^2).$$

We have $\square$. 

12
We now consider the function $f$ in the prototype problem (4). Then, we have the following result.

**Theorem 5.3.** Suppose that $f(X) \equiv \|YZ - W\|_F^2$, where $W \in \mathbb{Q}^{m \times n}$, $Y \in \mathbb{Q}^{m \times r}$ and $Z \in \mathbb{Q}^{r \times n}$. Then,

\[
\frac{\partial^2}{\partial W^2} f(X) \Delta W = \Delta W, \quad (15)
\]

\[
\frac{\partial^2}{\partial Y \partial W} f(X) \Delta Y = -\Delta Y Z, \quad (16)
\]

\[
\frac{\partial^2}{\partial Z \partial W} f(X) \Delta Z = -Y \Delta Z, \quad (17)
\]

\[
\frac{\partial^2}{\partial W \partial Y} f(X) \Delta W = -\Delta W Z^*, \quad (18)
\]

\[
\frac{\partial^2}{\partial Y^2} f(X) \Delta Y = \Delta Y ZZ^*, \quad (19)
\]

\[
\frac{\partial^2}{\partial Z \partial Y} f(X) \Delta Z = (YZ - X)(\Delta Z)^* + Y(\Delta Z)Z^*, \quad (20)
\]

\[
\frac{\partial^2}{\partial W \partial Z} f(X) \Delta W = -Y^* \Delta W, \quad (21)
\]

\[
\frac{\partial^2}{\partial Y \partial Z} f(X) \Delta Y = (\Delta Y)^*(YZ - X) + Y^*(\Delta Y)Z, \quad (22)
\]

\[
\frac{\partial^2}{\partial Y^2} f(X) \Delta Y = YY^* \Delta Z. \quad (23)
\]

**Proof.** By (9), we have

\[
\frac{\partial}{\partial W} f(W, Y, Z) = W - YZ.
\]

Thus,

\[
\frac{\partial}{\partial W} f(W + \Delta W, Y, Z) = W + \Delta W - YZ,
\]

\[
\frac{\partial}{\partial W} f(W + \Delta W, Y, Z) - \frac{\partial}{\partial W} f(W, Y, Z) = \Delta W.
\]

This implies (15). By above, we also have

\[
\frac{\partial}{\partial W} f(W, Y + \Delta Y, Z) = W - (Y + \Delta Y)Z,
\]

\[
\frac{\partial}{\partial W} f(W, Y + \Delta Y, Z) - \frac{\partial}{\partial W} f(W, Y, Z) = -\Delta Y Z.
\]

This implies (16). The formulas (17-23) can be proved similarly. \qed
We study the second order optimality conditions of (3) now.

Suppose that \( X^\# \in \mathbb{H} \) is a stationary point of (3), i.e., there exist some Langrangian multipliers \( \lambda_j, \mu_k \in \mathbb{R} \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \), such that (6), (7) and (8) are satisfied. Let

\[
I(X^\#) = \{k = 1, \cdots, p : g_k(X^\#) = 0\}.
\]

We say that \( \Delta X \in \mathbb{H} \) is in the critical cone \( C(X^\#) \) of (3) at \( X^\# \), if

\[
\nabla f(X^\#) \cdot \Delta X = 0,
\]

\[
\nabla h_j(X^\#) \cdot \Delta X = 0, \text{ for } j = 1, \cdots, p,
\]

\[
\nabla g_k(X^\#) \cdot \Delta X = 0, \text{ for } k \in I(X^\#).
\]

\[\tag{24}\]

\[\tag{25}\]

\[\tag{26}\]

We have the following theorems.

**Theorem 5.4.** Suppose that the functions \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) satisfy the strong assumption. Assume that \( X^\# = (W^\#, Y^\#, Z^\#) \in \mathbb{H} \) is an optimal solution of (3), and (5) is \( \mathbb{R} \)-linearly independent. Then for any \( \Delta X \in \mathbb{H} \),

\[
\frac{1}{2} \nabla^2 f(X^\#) \Delta X \cdot \Delta X \geq 0.
\]

**Theorem 5.5.** Suppose that the functions \( f, h_j \) and \( g_k \) for \( j = 1, \cdots, p \) and \( k = 1, \cdots, q \) satisfy the strong assumption. Assume that \( X^\# = (W^\#, Y^\#, Z^\#) \in \mathbb{H} \) is a stationary point of (3), and for any \( \Delta X \in \mathbb{H} \),

\[
\frac{1}{2} \nabla^2 f(X^\#) \Delta X \cdot \Delta X > 0.
\]

Then \( X^\# \) is an optimal minimizer of (3).

By regarding \( f, h_j, g_k \) for \( j = 1, \cdots, p \), \( k = 1, \cdots, q \) as functions of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \), from the second order optimality necessary condition and sufficient condition for real constrained optimization, we have Theorems 5.4 and 5.5.

### 6 Convex Functions of Quaternion Matrix Variables

Jia, Ng and Song [8] introduced subgradients of norms of quaternion matrix variables. Thus, they studied convex functions of quaternion matrix variables.

Suppose that \( f(X) : \mathbb{H} \to \mathbb{R} \). A natural definition for \( f \) to be convex is as follows. We say that \( f \) is a convex function if for any \( X, \hat{X} \in \mathbb{H} \), and any \( t \in \mathbb{R}, 0 \leq t \leq 1 \), we have

\[
f(tX + (1 - t)\hat{X}) \leq tf(X) + (1 - t)f(\hat{X}).
\]

It is possible to use the modern convex analysis terminology, epigraphs, to define convex functions of quaternion matrix variables. Here, we use the classical definition to make the definition, such that it is more convenient for engineering readers.
Then, a question is: Is \( f \) a convex function if and only if \( f \) is convex when \( f \) is regarded as a function of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \)?

**Proposition 6.1.** Suppose that \( f(\mathbf{X}) : \mathbb{H} \rightarrow \mathbb{R} \). Then \( f \) is a convex function if and only if \( f \) is convex when \( f \) is regarded as a function of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \).

**Proof.** Let \( \mathbf{X} = (W, Y, Z), \hat{\mathbf{X}} = \left( \hat{W}, \hat{Y}, \hat{Z} \right) \in \mathbb{H}, t \in \mathbb{R} \) and \( 0 \leq t \leq 1 \). Then

\[
t\mathbf{X} + (1 - t)\hat{\mathbf{X}} = \left( tW + (1 - t)\hat{W}, tY + (1 - t)\hat{Y}, tZ + (1 - t)\hat{Z} \right),
\]

\[
tW + (1 - t)\hat{W} = \left( tW_0 + (1 - t)\hat{W}_0 \right) + \left( tW_1 + (1 - t)\hat{W}_1 \right) \mathbf{i} + \left( tW_2 + (1 - t)\hat{W}_2 \right) \mathbf{j} + \left( tW_3 + (1 - t)\hat{W}_3 \right) \mathbf{k},
\]

\[
tY + (1 - t)\hat{Y} = \left( tY_0 + (1 - t)\hat{Y}_0 \right) + \left( tY_1 + (1 - t)\hat{Y}_1 \right) \mathbf{i} + \left( tY_2 + (1 - t)\hat{Y}_2 \right) \mathbf{j} + \left( tY_3 + (1 - t)\hat{Y}_3 \right) \mathbf{k},
\]

\[
tZ + (1 - t)\hat{Z} = \left( tZ_0 + (1 - t)\hat{Z}_0 \right) + \left( tZ_1 + (1 - t)\hat{Z}_1 \right) \mathbf{i} + \left( tZ_2 + (1 - t)\hat{Z}_2 \right) \mathbf{j} + \left( tZ_3 + (1 - t)\hat{Z}_3 \right) \mathbf{k}.
\]

From here, we may conclude that \( f \) is a convex function if and only if \( f \) is convex when \( f \) is regarded as a function of \( W_i, Y_i \) and \( Z_i, i = 0, 1, 2, 3 \). \( \square \)

We now define subgradients and subdifferentials by R-product. Suppose that \( f(\mathbf{X}) : \mathbb{H} \rightarrow \mathbb{R} \), and \( \bar{\mathbf{X}} = (\bar{W}, \bar{Y}, \bar{Z}) \in \mathbb{H} \). Let \( \mathbf{G} = (A, B, C) \in \mathbb{H} \). We say that \( \mathbf{G} \) is a subgradient of \( f \) at \( \bar{\mathbf{X}} \) if for any \( \mathbf{X} = (W, Y, Z) \in \mathbb{H} \), we have

\[
f(\mathbf{X}) \geq f(\bar{\mathbf{X}}) + \mathbf{G} : (\mathbf{X} - \bar{\mathbf{X}}).
\]

The set of all subgradients of \( f \) at \( \bar{\mathbf{X}} \) is called the subdifferential of \( f \) at \( \bar{\mathbf{X}} \) and denoted as \( \partial f(\bar{\mathbf{X}}) \).

By the definition of R-product, we see that \( \mathbf{G} \) is a subgradient of \( f \) at \( \bar{\mathbf{X}} \) if and only if \( R(\mathbf{G}) \) is a subgradient of \( f^R \) at \( R(\bar{\mathbf{X}}) \).

This definition is slightly different from the definition of Jia, Ng and Song \cite{8} in face. They used the real part of the inner product of \( \mathbf{G} \) and \( \mathbf{X} - \bar{\mathbf{X}} \), instead of their R-product here. By Proposition 4.1, these two definitions are the same. This also reveals that the subgradient concept introduced in \cite{8} can be regarded as subgradients of the norms of real matrix variables.

From Proposition 6.1, the definition of R-product and the knowledge of convex functions of real variables, we have the following proposition.
Proposition 6.2. Suppose that $f(X) : \mathbb{H} \rightarrow \mathbb{R}$ is a convex function. Then for any $\bar{X} = (\bar{W}, \bar{Y}, \bar{Z}) \in \mathbb{H}$, the subdifferential $\partial f(\bar{X})$ is a nonempty, convex and compact set in $\mathbb{H}$. The subdifferential $\partial f(\bar{X})$ is a singleton if and only if $f$ is differentiable at $\bar{X}$. In this case, $\nabla f(\bar{X})$ is the unique subgradient of $f$ at $\bar{X}$.

By Proposition 6.1 and the knowledge of convex functions of real variables, we have the following proposition.

Proposition 6.3. Suppose that $f(X) : \mathbb{H} \rightarrow \mathbb{R}$ satisfies the strong assumption. Then $f$ is convex if and only if $\nabla^2 f$ is positive semi-definite at any $X \in \mathbb{H}$.

7 A Sparse Color Image Denoising Model

We now consider a sparse color image denoising (SCID) model

$$\min \left\{ f(X) \equiv \frac{1}{2} \|L(Y + Z) - D\|_F^2 + \lambda \|Z\|_0 : \text{rank}(Y) \leq r \right\}, \quad (27)$$

Here, $D \in \mathbb{Q}^{m \times n}$ is an observation quaternion matrix of the color image, $Y \in \mathbb{Q}^{m \times n}$ is a low rank quaternion matrix to approximate $D$, $r$ is a prescribed integer for the upper bound of the rank of $Y$, $Z \in \mathbb{Q}^{m \times n}$ is the color image noise to be detected, $\lambda > 0$ is a prescribed parameter, $L : \mathbb{Q}^{m \times n} \rightarrow \mathbb{Q}^{m \times n}$ is a linear operator, for example, a projection operator to indicate the observed area, $X = (Y, Z)$ is the quaternion matrix vector variable, $f : \mathbb{M}^{m \times n} \equiv \mathbb{Q}^{m \times n} \times \mathbb{Q}^{m \times n} \rightarrow \mathbb{R}$ is the objective function. Assume that

$$Y = Y_0 + Y_1i + Y_2j + Y_3k, \quad Z = Z_0 + Z_1i + Z_2j + Z_3k, \quad D = D_0 + D_1i + D_2j + D_3k.$$

The task of this section is to analyze the optimality condition of (27) to pave the way for further study on similar color image models. As the 0-norm and the rank function are not continuous, we have to develop general subdifferential calculus for real lower semi-continuous functions of quaternion matrix variables first.

7.1 Generalized Subdifferentials of Lower Semi-Continuous Functions of Quaternion Matrix Variables

The generalized subdifferential calculus for real functions of real variables can be found in standard references [5, 13, 14]. In this subsection, we extend it to real functions of quaternion matrix variables.

Denote $\hat{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$. For a function $h(X) : \mathbb{M} \rightarrow \hat{\mathbb{R}}$, denote its domain as $\text{dom}(h) = \{X \in \mathbb{M} : h(X) < +\infty\}$. We say that $h$ is a proper function if its domain is not empty.
Suppose that $h(\mathbf{X}) : \mathbb{M} \to \bar{\mathbb{R}}$ is a proper function, $\bar{\mathbf{X}} = (\bar{Y}, \bar{Z}) \in \text{dom}(h)$. Let $\mathbf{G} = (A, B) \in \mathbb{M}$. We say that $\mathbf{G}$ is a F(réchet)-subgradient of $h$ at $\bar{\mathbf{X}}$ if
\[
\liminf_{\mathbf{X} \to \bar{\mathbf{X}}, \mathbf{X} \neq \bar{\mathbf{X}}} \frac{h(\mathbf{X}) - h(\bar{\mathbf{X}}) - \mathbf{G} \cdot (\mathbf{X} - \bar{\mathbf{X}})}{\|\mathbf{X} - \bar{\mathbf{X}}\|} \geq 0.
\]
The set of all F-subgradients of $h$ at $\bar{\mathbf{X}}$ is called the Fréchet subdifferential of $h$ at $\bar{\mathbf{X}}$ and denoted as $\partial^F h(\bar{\mathbf{X}})$. On the other hand, $\mathbf{G} \in \mathbb{M}$ is a limiting subgradient of $h$ at $\bar{\mathbf{X}}$ if
\[
\mathbf{G} = \lim_{k \to \infty} \mathbf{G}^k, \quad \mathbf{G}^k \in \partial^F h(\mathbf{X}^k), \quad \lim_{k \to \infty} \mathbf{X}^k = \mathbf{X}.
\]
The set of all limiting subgradients of $h$ at $\bar{\mathbf{X}}$ is called the limiting subdifferential of $h$ at $\bar{\mathbf{X}}$ and denoted as $\partial^L h(\bar{\mathbf{X}})$.

By the definition of R-product, we see that $\mathbf{G}$ is a F-subgradient (limiting subgradient) of $h$ at $\bar{\mathbf{X}}$ if and only if $R(\mathbf{G})$ is a F-subgradient (limiting subgradient) of $h^R$ at $R(\bar{\mathbf{X}})$. Here, $h$ is regarded as a function of $Y_i$ and $Z_i$ for $i = 0, 1, 2, 3$, and denote such a function as $h^R$, and for $\mathbf{X} = (Y, Z)$, denote $R(Y) = (Y_0, Y_1, Y_2, Y_3)$, $R(Z) = (Z_0, Z_1, Z_2, Z_3)$ and $R(\mathbf{X}) = (R(Y), R(Z))$.

By (3) of [9], we have
\[
\partial^F h(\bar{\mathbf{X}}) \subset \partial^L h(\bar{\mathbf{X}}).
\]

By Theorem 1 of [9], we have the following theorem.

**Theorem 7.1.** Let $A = (a_{ij}) \in \mathbb{Q}^{m \times n}$. Then
\[
\partial^F \|A\|_0 = \partial^L \|A\|_0 = \mathbb{Q}^{m \times n}_{\Gamma^c_A},
\]
where $\mathbb{Q}^{m \times n}_{\Gamma^c_A} = \{B \in \mathbb{Q}^{m \times n} : B_{\Gamma_A} = O\}, \Gamma_A = \{(i, j) : a_{ij} \neq 0\}$ is the support of $A$, and $\Gamma^c_A$ is the complementary set of $\Gamma_A$.

By using Theorem 4 of [9], we may also characterize the generalized subdifferential of the rank function of $A$.

### 7.2 The Feasibility Set of The SCID Optimization Problem

In this subsection, we study the feasibility set of the SCID optimization problem (27):
\[
S = \{Y \in \mathbb{Q}^{m \times n} : \text{rank}(Y) \leq r\}.
\]
Let
\[
R(S) = \{R(Y) : Y \in S\}.
\]
In this paper, we say that $R(S)$ is the real representation set of $S$. Note that it is different from $\{Y^R : Y \in S\}$.

**Proposition 7.2.** The set $S$ is closed. The set $R(S)$ is a closed and semi-algebraic set.
Proof. Suppose that $Y^k \in S$ and $Y^k \to Y$. By Theorem 2.1,
\[ Y^k = U^k \left( \frac{\Sigma^k}{O} \right) (V^k)^*, \]
where $U^k \in \mathbb{Q}^{m \times m}$ and $V^k \in \mathbb{Q}^{n \times n}$ are unitary, and $\Sigma^k = \text{diag}\{\sigma_{1,k}, \ldots, \sigma_{r,k}\}$ is a real nonnegative $r \times r$ diagonal matrix. Since $U^k$ and $V^k$ are unitary, $\{U^k\}$ and $\{V^k\}$ are bounded. Thus, $\{U^k\}$ and $\{V^k\}$ have limiting points. Without loss of generality, we may assume that $U^k \to U$ and $V^k \to V$. Then $U$ and $V$ are unitary. Since the rank function of quaternion matrices is lower semi-continuous, without loss of generality, we may also assume that $\sigma_{i,k} \to \sigma_i$ for $i = 1, \ldots, r$. Then we see that $\text{rank}(Y) \leq r$. This shows that $S$ is closed.

Since $S$ is closed, the set $R(S)$ is also closed. By Theorem 1.8.4 of [20], if the rank of $A \in \mathbb{Q}^{m \times n}$ is $k$, then the rank of its real representation $A^R$ is $4k$. Note that
\[ R(S) = \left\{ R(Y) : Y \in \mathbb{Q}^{m \times n} : \text{rank}(Y) \leq r \right\} \]
\[ = \left\{ R(Y) : Y \in \mathbb{Q}^{m \times n} : \text{rank}(Y^R) \leq 4r \right\}. \]
As the low-rank real matrix set is semi-algebraic [6], we conclude that the set $R(S)$ is semi-algebraic.

For a nonempty closed set $\Omega$, the indicator function with respect to $\Omega$, denoted as $\delta_\Omega$ is defined by
\[ \delta_\Omega(x) = \begin{cases} 0, & \text{if } x \in \Omega; \\ +\infty, & \text{otherwise}. \end{cases} \]

Let $A \in \mathbb{Q}^{m \times n}$ and $B \to A$. Then there is $\delta > 0$ such that for all $B \in \mathbb{Q}^{m \times n}$, satisfying $\|B - A\|_F \leq \delta$, we have $\|B\|_0 \geq \|A\|$. This shows that the 0-norm function of quaternion matrices is lower semi-continuous.

**Proposition 7.3.** Suppose that $p : \mathbb{M} \to \mathbb{R} \cup \{+\infty\}$ is defined by $p(X) = \lambda\|Z\|_0 + \delta_S(Y)$, where $X = (Y, Z)$. Let $p^R(R(Y), R(Z)) \equiv p(X)$. Then,

(i) $p$ is a proper lower semi-continuous function, and $p^R$ is a semi-algebraic function;
(ii) the limiting subdifferential of $h$ takes the form of
\[ \partial^L p(Y, Z) = N_S(Y) \times \mathbb{Q}^{m \times n}_{\Gamma^2}, \]
where $N_S(Y)$ is the normal cone with respect to $S$ at $Y$.

**Proof.** (i) For any $Y \in S$ and $Z \in \mathbb{Q}^{m \times n}$, $p(X)$ is nonnegative and finite valued. Thus, $p$ is proper. Since the 0-norm function of quaternion matrices is lower semi-continuous, and $S$ is closed by Proposition 7.2, $p$ is lower semi-continuous. By Proposition 7.2, $R(S)$ is semi-algebraic. The 0-norm function has a piecewise linear graph, see [1]. Thus, $p^R$ is a semi-algebraic function.
(ii) Note that $p$ is separable in $Y$ and $Z$. Thus,

$$
\partial^L p(Y, Z) = \partial^L \delta_S(Y) \times \partial^L (\lambda\|Z\|_0) = N_S(Y) \times \mathbb{Q}_{\Gamma}^{m\times n},
$$

where the second equality comes from Theorem 7.1. \qed

The semi-algebraic property of $p^R$ will not be used in this paper, yet it is very important in convergence analysis of first-order algorithms for solving (27) [1].

### 7.3 Stationarity

The SCID model (27) is a rank-constrained sparse optimization problem of quaternion matrix variables. By sparse optimization [2] and rank-constrained optimization [11], we may introduce stationary point concepts for (27).

We first introduce proximal mapping for a lower semi-continuous function $h : \mathbb{M} \rightarrow \mathbb{R} \cup \{+\infty\}$. We denote is as Prox$_h$. It is defined as

$$
\text{Prox}_h(X) := \arg \min_{W \in \mathbb{M}} \left\{ h(W) + \frac{1}{2} \|W - X\|^2_F \right\}.
$$

Then we define $\beta$–stationary points and stationary points for (27). Let

$$
h(X) = \frac{1}{2}\|L(Y + Z) - D\|^2_F.
$$

Let $\beta$ be a positive number, $\bar{X} = (\bar{Y}, \bar{Z}) \in \mathbb{M}$. We say that $\bar{X}$ a $\beta$–stationary point of (27) if

$$
\bar{Y} \in \Pi_S(\bar{Y} - \beta \nabla_Y h(\bar{X})), \quad \bar{Z} \in \text{Prox}_{\beta \lambda\|\cdot\|_0}(\bar{Y} - \beta \nabla_Z h(\bar{X}));
$$

we say that $\bar{X}$ a stationary point of (27) if

$$
\nabla_Y h(\bar{X}) \in N_S(\bar{Y}), \quad \nabla_Z h(X) \in \mathbb{Q}_{\Gamma}^{m\times n}.
$$

By (11), we have

$$
\nabla h(X) = \left[ \frac{\partial}{\partial Y} h(X), \frac{\partial}{\partial Z} h(X) \right] = [L^* (L(Y + Z) - D), L^* (L(Y + Z) - D)].
$$

**Proposition 7.4.** There is a gradient Lipschitz constant $L_h = \sqrt{2\|L^*L\|_F} > 0$ such that for any $X = (Y, Z), \bar{X} = (\bar{Y}, \bar{Z}) \in \mathbb{M},$

$$
\|\nabla h(X) - \nabla h(\bar{X})\| \leq L_h \|X - \bar{X}\|_F.
$$

**Proof.** By (30), we have the conclusion. \qed
Theorem 7.5. For the SCID model (27), we have the following conclusions:

(i) Any local minimizer is a stationary point;
(ii) Any $\beta-$stationary point for $\beta > 0$ is a stationary point;
(iii) Any global minimizer $X^* = (Y^*, Z^*)$ is a $\beta-$stationary point for $\beta \in (0, \frac{1}{L_h})$; furthermore, $\Pi_S(Y^* - \beta \nabla_Y f(X^*))$ and $\text{Prox}_{\beta \|\cdot\|_0}(Z^* - \beta \nabla_Z f(X^*))$ are singletons;
(iv) If $X^* = (Y^*, Z^*)$ is a $\beta-$stationary point and rank($Y^*$) < $r$, then $X^*$ is a local minimizer.

Proof. (i) Let $\bar{X} = (\bar{Y}, \bar{Z})$ be a local minimizer of (27). Rewrite problem (27) as

$$\min_X h(X) + p(X), \quad (31)$$

where $h$ is defined by (28), $p$ is defined in Proposition 7.3. Note that the quaternion matrix optimization problem (31) is equivalent to the real matrix optimization problem

$$\min_{R(X)} h^R(R(X)) + p^R(R(X)). \quad (32)$$

Apply the generalized Fermat rule [14, Theorem 10.1] to the real matrix optimization problem (32). Since $G$ is a F-subgradient (limiting subgradient) of $h$ ($p$) at $\bar{X}$ if and only if $R(G)$ is a F-subgradient (limiting subgradient) of $h^R$ ($p^R$) at $R(\bar{X})$, we have the desired result.

(ii) Let $\bar{X} = (\bar{Y}, \bar{Z})$ be a $\beta-$stationary point of (27). Note that

$$\bar{Y} \in \Pi_S(\bar{Y} - \beta \nabla_Y h(\bar{X}))$$

is equivalent to

$$R(\bar{Y}) \in \Pi_{R(S)}(R(\bar{Y}) - \beta \nabla_{R(Y)} h^R(R(\bar{X}))),$$

which is further equivalent to

$$R(\bar{Y}) = \arg\min_{R(Y)} \left\{ \frac{1}{2} \|R(Y) - (R(\bar{Y}) - \beta \nabla_{R(Y)} h^R(R(\bar{X})))\|_F^2 + \delta_{R(S)}(Y) \right\}. \quad (33)$$

Combining this with the generalized Fermat rule [14, Theorem 10.1] and the fact that

$$\partial_{R(y)} \delta_{R(S)}(R(Y)) = N_{R(S)}(R(Y))$$

[14 Exercise 8.14], we conclude that

$$O \in R(\bar{Y}) - \beta \nabla_{R(Y)} h^R(R(\bar{X}) + N_{R(S)}(\bar{Y}),$$

which is exactly

$$\nabla_{R(Y)} h^R(R(X)) \in N_{R(S)}(R(\bar{Y})), \quad (33)$$

because of the conic property of $N_{R(S)}(R(\bar{Y}))$. Note that (33) is equivalent to

$$\nabla_Y h(X) \in N_S(\bar{Y}). \quad (34)$$
On the other hand, the relation

\[ \tilde{Z} \in \text{Prox}_{\beta\lambda \| \cdot \|_0} (\tilde{Y} - \beta \nabla Z h(\tilde{X})) \]

indicates that

\[ \nabla Z h(\tilde{X}) \in \mathbb{Q}_{m \times n}^m. \]

Combining (34), we have Conclusion (ii).

(iii)

(iv) Since the Eckart-Young-Mirsky low-rank approximation theorem can be applied to the case of quaternion matrices [8], we can easily verify the implication below:

\[ Y \in \Pi_S(Y + T) \ & \ \text{rank}(Y) < r \implies T = O. \] (35)

If \( X^* = (Y^*, Z^*) \) is a \( \beta \)-stationary point with \( \text{rank}(Y^*) < r \), then by the definition of \( \beta \)-stationarity in (29), together with (30), we have \( \nabla_Y h(X^*) = \nabla_Z h(X^*) = O \) by employing (35). Additionally, the convexity of \( h \) yields that for any \( X \in \mathbb{Q}_{m \times n}^m \),

\[ h(X) \geq h(X^*) + \nabla h(X^*) \cdot (X - X^*) = h(X^*). \] (36)

Note that \( \|Z\|_0 \) only takes integer values \( 0, 1, \ldots, mn \). Thus we can find some positive scalar \( \epsilon \) such that

\[ \|Z\|_0 \geq \|Z^*\|_0, \ \forall Z \in N(Z^*, \epsilon), \]

where \( N(Z^*, \epsilon) := \{Z : \|Z - Z^*\|_F \leq \epsilon\} \). Combining with (36), we can conclude that for any feasible solution \( X \in N(X^*, \epsilon) \), \( h(X) + p(X) \geq h(X^*) + p(X^*) \), that is, \( X^* \) is a local minimizer. This completes the proof. \( \square \)

8 Final Remarks

In this paper, we introduce first and second order derivatives of real functions of quaternion matrix variables, and established their calculation rules. Our approach is consistent with the subgradient concept for norms of quaternion matrix variables introduced in [8]. We established first and second order optimality conditions for constrained optimization problems of real functions in quaternion matrix variables. Optimization methods can be developed based upon these.

One key tool of our approach is the R-product. It turns out that the R-product of two quaternion matrices is equal to the real part of the inner product of these two quaternion matrices. This is not by chance. As we may form a third order real tensor to a quaternion matrix, there is also an inverse reaction to make the operation results of quaternion matrices to the real field, such as singular values and R-products.

Finally, we introduce the generalized subdifferentials of lower semi-continuous functions of quaternion matrices, and use them to analyze the optimality conditions of a sparse low rank
color image denoising model. This combines the knowledge of quaternion matrices, color image processing and variational analysis.

We hope that our work is useful to people working with optimization models involving quaternion matrices.

Acknowledgment We are thankful to Prof. Michael Ng for the discussion on subgradients of norms of quaternion matrices, and to Prof. Defeng Sun for the discussion on second order derivatives of real functions with quaternion matrix variables.

References

[1] H. Attouch, J. Bolte and B.F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods”, Math. Prog. 137 (2013) 91–129.

[2] A. Beck and Y.C. Eldar, “Sparsity constrained nonlinear optimization: Optimality conditions and algorithms”, SIAM J. Optim. 23 (2013) 1480-1509.

[3] Y. Chen, L. Qi, X. Zhang and Y. Xu, “A low rank quaternion decomposition algorithm and its application in color image inpainting”, arXiv:2009.12203 (2020).

[4] Y. Chen, X. Xiao and Y. Zhou, “Low-rank quaternion approximation for color image processing”, IEEE Transactions on Image Processing 29 (2019) 1057-7149.

[5] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, 1983.

[6] J-B. Hiriart-Urruty and H.Y. Le, “From Eckart and Young approximation to Moreau envelopes and vice versa”, RAIRO Operations Research 47 (2013) 299-310.

[7] J-B. Hiriart-Urruty and H.Y. Le, “A variational approach of the rank function”, TOP 21 (2013) 207-240.

[8] Z. Jia, M.K. Ng and G. Song, “Robust quaternion matrix completion with applications to image inpainting”, Numerical Linear Algebra with Applications 26 (2019) e2245.

[9] H.Y. Le, “Generalized subdifferentials of the rank function”, Optim. Lett. 7 (2013) 731-743.

[10] N. Le Bihan and S.J. Sangwine, “Quaternion principal component analysis of color images”, IEEE International Conference of Image Process 1 (2003) 809-812.

[11] X. Li, N. Xie and W. Song, “Optimality conditions for rank-constrained matrix optimization”, J. Oper. Res. Soc. China 7 (2019) 285-301.
[12] D.P. Mandic, C. Jahanchahi, C.C. Took, “A quaternion gradient operator and its applications”, *IEEE Signal Processing Letter* 18 (2011) 47-50.

[13] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basis Theory*, Springer, 2006.

[14] T.R. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer, 2009.

[15] S.J. Sangwine, “Fourier transforms of colour images using quaternion or hypercomplex, numbers”, *Electronics Letters* 32 (1996) 1979-1980

[16] S.J. Sangwine and N. Le Bihan, “Quaternion singular value decomposition based on bidiagonalization to a real or complex matrix using quaternion Householder transformations”, *Applied Mathematics and Computation* 182 (2006) 727-738.

[17] X. Xiao and Y. Zhou, “Two-dimensional quaternion pca and sparse pca”, *IEEE Trans. Neural Netw. Lear. Syst.* (2018).

[18] D. Xu, C. Jahanchahi, C.C. Took and D.P. Mandic, “Enabling quaternion derivatives: the generalized HR calculus”, *R. Soc. open sci.* 2 (2015) 150255.

[19] Y. Xu, L. Yu, H. Xu, H. Zhang and T. Nguyen, “Vector sparse representation of color image using quaternion matrix analysis”, *IEEE Trans. Image Process.* 24 (2015) 1315-1329.

[20] M. Wei, Y. Li, F. Zhang and J. Zhao, *Quaternion Matrix Computations*, Nova Science Publisher, New York, 2018.

[21] F. Zhang, “Quaternions and matrices of quaternions”, *Linear Algebra and Its Applications* 251 (1997) 21-57.

[22] C. Zou, K.I. Kou and Y. Wang, “Quarernion collaborative and sparse representation with application to color face recognition”, *IEEE Trans. Image Process.* 25 (2016) 3287-3302.