Quasi Non-Negative Quaternion Matrix Factorization with Application to Color Face Recognition

Yifen Ke\textsuperscript{1,2} · Changfeng Ma\textsuperscript{1,2} · Zhigang Jia\textsuperscript{3,4} · Yajun Xie\textsuperscript{5} · Riwei Liao\textsuperscript{6}

Received: 16 August 2022 / Revised: 13 February 2023 / Accepted: 13 February 2023 / Published online: 16 March 2023

Abstract
To address the non-negativity dropout problem of quaternion models, a novel quasi non-negative quaternion matrix factorization (QNQMF) model is presented for color image processing. To implement QNQMF, the quaternion projected gradient algorithm and the quaternion alternating direction method of multipliers are proposed via formulating QNQMF as the non-convex constraint quaternion optimization problems. Some properties of the proposed algorithms are studied. The numerical experiments on the color image reconstruction show that these algorithms encoded on the quaternion perform better than these algorithms encoded on the red, green and blue channels. Furthermore, we apply the proposed algorithms to the color face recognition. Numerical results indicate that the accuracy rate of face recognition on the quaternion model is better than on the red, green and blue channels of color image as well as single channel of gray level images for the same data, when large facial expressions and shooting angle variations are presented.

Keywords Quaternion matrix · Quasi non-negative quaternion matrix factorization · Quaternion optimization · Color face recognition

Mathematics Subject Classification 15B33 · 90C26

1 Introduction
Color plays an important role in image processing tasks, especially in complicated scenes. The color principal component analysis [21] preserves the messages between channels as well as the important low-frequency information. However, it is expensive to compute the dominant eigenvectors of quaternion covariance matrix of large-scale sizes and the non-negativity of pixel values are not well preserved. In this paper, we present a new quasi non-negative quaternion matrix factorization model with non-negative constraints on the imaginary parts of quaternion matrix factors, and propose two solvers: the quaternion projected gradient...
The proposed model and numerical algorithms are successfully applied to the color face recognition, and numerical results demonstrate its efficiency and robustness, particularly in large-scale size faces, large facial expressions and shooting angle variations.

For given color and gray level images, the objects in color images can be recognized easier and faster than their counterparts in gray images. Recently, it has been evinced that color does contribute to face recognizing especially when shape cues of the images are degraded [41]. In [33], Rajapakse, Tan and Rajapakse used the non-negative matrix factorization (NMF) model to the color face recognizing, which encoded separately the red, green, blue channels and projected together these feature vectors to sparse subspaces. The numerical results presented in [33] show that the improved accuracy of color image face recognition over gray image face recognition while large facial expressions and illumination variations are appeared. For the color images, it is a common approach to treat separately for the red, green and blue channels, and then the final processed image is obtained by combining the results from the color channels. The downside of this approach is that the messages between channels are often ignored. It is obvious that we can use a third-order tensor to represent the red, green and blue channels for the color image [39]. Each slice of this third-order tensor corresponds to one channel of the color image. In general, it can consider the non-negative tensor factorization (NTF) to complete the low-rank and non-negative approximation. However, the theory for NTF is not well-established when it comes to the tensor rank and inverse for the general tensor.

Recently, the quaternion matrix representation is playing an increasingly important role in the color image processing problems. Pure quaternion matrices are used to represent color images, that is, embedding the red, green and blue channel matrices into three imaginary parts and keeping the real part zero. Based on this, there are many successful models, such as the quaternion principal component analysis for color face recognition [38, 41], the quaternion Schur decomposition for color watermarking [5, 25], the quaternion singular value decomposition for color image inpainting [6, 19, 20], and so on. For more details about quaternion matrix and its applications, we refer to [15, 16, 18, 22, 27, 35, 37] and the references therein. Unfortunately, the above quaternion decomposition models ignore a fundamental property, which is the non-negativity of the imaginary parts that representing color channels.

To address this long-standing non-negativity dropout problem of quaternion models, we propose a quasi non-negative quaternion matrix factorization (QNQMF) model. This factorization possesses the simple form, good interpretability, small storage space, and takes full advantage of color information between channels for color images. In order to implement this decomposition, we formulate QNQMF as a quaternion optimization problem. The most difficult thing is that the obtained quaternion optimization problem is a multi-variable coupled, non-convex and non-linear complicated optimization problem. We propose the quaternion projected gradient algorithm and the quaternion alternating direction method of multipliers to solve QNQMF. Then this factorization is applied to the color face recognition. The improved accuracy of color image recognition indicates the recognition performance and robustness of the implementation on the QNQMF model for color face images.

Recall that the non-negative matrix factorization (NMF) is proposed as a new matrix factorization model by Lee and Seung [24] in 1999, and has been developed well in both theory and application. Due to the simple form, good interpretability, small storage space as well as other advantages, NMF as a powerful tool for data analysis has been successfully used to data mining, pattern recognition and machine learning community [24, 30]. NMF problem is usually formulated as a non-convex and non-linear constraint optimization problem. On account of the practical application problems involved in NMF, there exists many variants of
NMF using other equivalent objective functions and additional constraints on the low-rank and non-negative factors (e.g., sparsity, orthogonality and smoothness); see [28, 31, 34]. Many numerical algorithms have been proposed to solve NMF, for example, the multiplicative update method [24], the alternating non-negative least squares method [1, 7], the Newton-like method [11], the projected gradient method [26], the active set method [23], the alternating direction method of multipliers [13, 40], the NeNMF method [12] and so on. As far as we know, NMF has not been extended to quaternion matrices with requiring the non-negative constraints on the imaginary parts of quaternion matrix until now.

This paper is organized as follows. In Sect. 2, we review the basic concepts of quaternion matrices and non-negative (real) matrix factorization. In Sect. 3, a new QNQMF model is introduced via formulating nonion matrices and non-negative (real) matrix factorization. In Sect. 3, a new QNQMF model is introduced via formulating nonion matrices and non-negative (real) matrix factorization. In Sect. 6, we offer the general conclusions of this work in Sect. 6.

Throughout this paper, we denote the real field and the quaternion algebra by $\mathbb{R}$ and $\mathbb{Q}$, respectively. The sets of real and quaternion $m \times n$ matrices are denoted by $\mathbb{R}^{m \times n}$ and $\mathbb{Q}^{m \times n}$, respectively. Let $\mathbb{R}_{+}^{m \times n}$ be the non-negative set on the real field. We use the black letters $A, B, C, \ldots$ to denote the quaternion matrices and the letters $a, b, c, \ldots$ to denote the real matrices. The symbol $\odot$ denotes the Hadamard product, that is $(A \odot B)_{st} = a_{st}b_{st}$ for the same size real matrices $A = (a_{st})$ and $B = (b_{st})$. The symbol $I$ is the identity matrix and $\| \cdot \|_2$ denotes the Euclidean norm of the real vector.

2 Preliminaries

In this section, we present several basic results of quaternion matrices and the non-negative matrix factorization on the real field.

Quaternions, introduced by the mathematician Hamilton [14] in 1843, are generalizations of the real and complex numbers systems. A quaternion $q$ is defined by

$$q = q_0 + q_1i + q_2j + q_3k,$$

where $q_0, q_1, q_2, q_3$ are real numbers, $i, j$ and $k$ are imaginary numbers satisfying

$$i^2 = j^2 = k^2 = -1, \quad ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j.$$

Let $p = p_0 + p_1i + p_2j + p_3k \in \mathbb{Q}$ and $q = q_0 + q_1i + q_2j + q_3k \in \mathbb{Q}$. The sum $p + q$ of $p$ and $q$ is

$$p + q = (p_0 + q_0) + (p_1 + q_1)i + (p_2 + q_2)j + (p_3 + q_3)k$$

and their multiplication $pq$ is

$$pq = (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)i$$
$$+ (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)j + (p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0)k.$$
The real part of $q$ is $\text{Re} \ q = q_0$, and its imaginary is $\text{Im} \ q = q_1 \mathbf{i} + q_2 \mathbf{j} + q_3 \mathbf{k}$. For the convenience of description, we denote $\text{Im}_1 \ q = q_1$, $\text{Im}_j \ q = q_2$ and $\text{Im}_k \ q = q_3$. When $\text{Re} \ q = 0$, a non-zero quaternion $q$ is said to be pure imaginary. The quaternion conjugate of $q$ is denoted by $\overline{q} = q^* = \text{Re} \ q - \text{Im} \ q = q_0 - q_1 \mathbf{i} - q_2 \mathbf{j} - q_3 \mathbf{k}$. The modulus of $q$ is defined by $|q| = \sqrt{\overline{q}q} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$. If $q \neq 0$, then $q^{-1} = \frac{\overline{q}}{|q|^2}$. We refer to [17] for more details about quaternions.

A quaternion matrix $A \in \mathbb{Q}^{m \times n}$ can be denoted as

$$A = A_0 + A_1 \mathbf{i} + A_2 \mathbf{j} + A_3 \mathbf{k}, \quad A_0, A_1, A_2, A_3 \in \mathbb{R}^{m \times n}.$$ 

The transpose of $A$ is $A^T = A_0^T + A_1^T \mathbf{i} + A_2^T \mathbf{j} + A_3^T \mathbf{k}$. The conjugate of $A$ is $\overline{A} = A_0 - A_1 \mathbf{i} - A_2 \mathbf{j} - A_3 \mathbf{k}$. The conjugate transpose of $A$ is $A^* = A_0^T - A_1^T \mathbf{i} - A_2^T \mathbf{j} - A_3^T \mathbf{k}$.

We remark that the identity quaternion matrix is the same as the classical identity matrix. Thus, we also denote as $I$. A quaternion matrix $A \in \mathbb{Q}^{n \times n}$ is inverse if there exists $B \in \mathbb{Q}^{n \times n}$ such that $AB = BA = I_n$. And the inverse matrix of $A$ is denoted by $A^{-1} = B$.

For $A = A_0 + A_1 \mathbf{i} + A_2 \mathbf{j} + A_3 \mathbf{k} = (a_{st}), \quad B = B_0 + B_1 \mathbf{i} + B_2 \mathbf{j} + B_3 \mathbf{k} = (b_{st}) \in \mathbb{Q}^{m \times n}$, their inner product is defined according to [20] by

$$\langle A, B \rangle = \text{Tr}(B^*A) = \sum_{s=1}^{m} \sum_{t=1}^{n} \overline{b}_{st} a_{st} \in \mathbb{Q},$$

where $\text{Tr}(B^*A)$ denotes the trace of $B^*A$. In addition, it holds that

$$\text{Re} \ \langle A, B \rangle = \text{Re} \ \langle B, A \rangle = \langle A_0, B_0 \rangle + \langle A_1, B_1 \rangle + \langle A_2, B_2 \rangle + \langle A_3, B_3 \rangle.$$ 

The Frobenius norm of $A = A_0 + A_1 \mathbf{i} + A_2 \mathbf{j} + A_3 \mathbf{k} = (a_{st}) \in \mathbb{Q}^{m \times n}$ is defined according to [20] by

$$\|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\text{Tr}(A^*A)} = \sqrt{\sum_{s=1}^{m} \sum_{t=1}^{n} |a_{st}|^2} \quad \text{(2.1)}$$

For a quaternion matrix $A = A_0 + A_1 \mathbf{i} + A_2 \mathbf{j} + A_3 \mathbf{k} \in \mathbb{Q}^{m \times n}$, its real representation is of the following form

$$\Upsilon_A = \begin{pmatrix} A_0 & A_2 & A_1 & A_3 \\ -A_2 & A_0 & A_3 & -A_1 \\ -A_1 & -A_3 & A_0 & A_2 \\ -A_3 & A_1 & -A_2 & A_0 \end{pmatrix} \in \mathbb{R}^{4m \times 4n}.$$ 

It is easy to verify that $\Upsilon_{AB^*} = \Upsilon_A (\Upsilon_B)^T$ for any $A \in \mathbb{Q}^{m \times l}$ and $B \in \mathbb{Q}^{n \times l}$. We refer to [17, 20] for more details about quaternion matrix.

At the end of this section, we recall the non-negative matrix factorization [24] on the real field. Recall that a real matrix $X$ is non-negative if and only if all entries of $X$ are non-negative, denoted by $X \geq 0$.

**Non-Negative Matrix Factorization (NMF) [24]**. For a given real matrix $X \in \mathbb{R}^{m \times n}$ with $X \geq 0$ and a pre-specified positive integer $l < \min(m, n)$, it finds two non-negative matrices $W \in \mathbb{R}^{m \times l}$ and $H \in \mathbb{R}^{l \times n}$ such that

$$X = WH. \quad \text{(2.2)}$$
3 Quasi Non-Negative Quaternion Matrix Factorization

In this section, a new quasi non-negative quaternion matrix factorization model is presented and its corresponding quaternion optimization problem is discussed.

**Definition 3.1** A quaternion matrix $A = A_0 + A_1i + A_2j + A_3k \in \mathbb{Q}^{m \times n}$ is called the quasi non-negative quaternion matrix if $A_1$, $A_2$ and $A_3$ are real non-negative matrices, that is

$$A_1 \geq 0, \quad A_2 \geq 0, \quad A_3 \geq 0.$$  

The set of quasi non-negative quaternion matrices is denoted by $\mathbb{Q}^{m \times n}_{\mathbb{Q}_n}$. 

**Quasi Non-Negative Quaternion Matrix Factorization (QNQMF).** For a given quaternion matrix $X = X_0 + X_1i + X_2j + X_3k \in \mathbb{Q}^{m \times n}_{\mathbb{Q}_n}$, it finds the matrices $W = W_0 + W_1i + W_2j + W_3k \in \mathbb{Q}^{l \times l}_{\mathbb{Q}_n}$ and $H = H_0 + H_1i + H_2j + H_3k \in \mathbb{Q}^{l \times n}_{\mathbb{Q}_n}$ such that

$$X = WH,$$  

that is

$$X_0 + X_1i + X_2j + X_3k$$

$$= (W_0H_0 - W_1H_1 - W_2H_2 - W_3H_3) + (W_0H_1 + W_1H_0 + W_2H_3 - W_3H_2)i$$

$$+ (W_0H_2 - W_1H_3 + W_2H_0 + W_3H_1)j + (W_0H_3 + W_1H_2 - W_2H_1 + W_3H_0)k.$$  

Here, $l$ is a pre-specified positive integer with $l < \min(m, n)$. The matrix $W$ is called the source matrix and the matrix $H$ is called the activation matrix.

We take a 4-by-4 quasi non-negative quaternion matrix for illustration. Let $X = X_0 + X_1i + X_2j + X_3k$ with

$$X_0 = \begin{pmatrix}
-6 & 3 & -2 & -9 \\
2 & 9 & 2 & -5 \\
-5 & 1 & -3 & -7 \\
-4 & 7 & 0 & -11
\end{pmatrix}, \quad X_1 = \begin{pmatrix}
3 & 3 & 7 & 3 \\
4 & 2 & 6 & 2 \\
0 & 2 & 4 & 0 \\
2 & 0 & 8 & 4
\end{pmatrix},$$

$$X_2 = \begin{pmatrix}
9 & 1 & 0 & 5 & 0 \\
8 & 4 & 2 & 4 \\
6 & 6 & 4 & 0 \\
14 & 12 & 8 & 4
\end{pmatrix}, \quad X_3 = \begin{pmatrix}
2 & 5 & 0 & 1 \\
4 & 3 & 4 & 5 \\
3 & 6 & 1 & 0 \\
2 & 7 & 2 & 1
\end{pmatrix}.$$  

Then there exist two quasi non-negative quaternion matrices $W = W_0 + W_1i + W_2j + W_3k \in \mathbb{Q}^{4 \times 1}_{\mathbb{Q}_n}$ and $H = H_0 + H_1i + H_2j + H_3k \in \mathbb{Q}^{1 \times 4}_{\mathbb{Q}_n}$ such that $X = WH$, in which

$$W_0 = \begin{pmatrix}
2 \\
3 \\
1 \\
3
\end{pmatrix}, \quad W_1 = \begin{pmatrix}
1 \\
0 \\
1 \\
0
\end{pmatrix}, \quad W_2 = \begin{pmatrix}
2 \\
0 \\
1 \\
2
\end{pmatrix}, \quad W_3 = \begin{pmatrix}
2 \\
1 \\
2 \\
3
\end{pmatrix}$$

and

$$H_0 = \begin{pmatrix}
1 & 3 & 1 & -1
\end{pmatrix}, \quad H_1 = \begin{pmatrix}
2 & 1 & 2 & 1
\end{pmatrix},$$

$$H_2 = \begin{pmatrix}
2 & 1 & 0 & 1
\end{pmatrix}, \quad H_3 = \begin{pmatrix}
1 & 0 & 1 & 2
\end{pmatrix}.$$  

Here, $X_0$, $W_0$ and $H_0$ can contain negative elements. Note that QNQMF only the three imaginary parts are required to be non-negative, which greatly enriches the range of feasible factors.
Remark 3.1 In [8], Flamant, Miron and Brie considered the following model:

\[ X = WH, \]

where \( X \in \mathbb{Q}_S^{m \times n}, \ W \in \mathbb{Q}_S^{m \times l} \) and \( H \in \mathbb{R}_+^{l \times n} \). Here, the set \( \mathbb{Q}_S^{m \times n} \) is defined as

\[ \mathbb{Q}_S^{m \times n} = \{ A = (a_{st}) \in \mathbb{Q}^{m \times n} \mid \text{Re}(a_{st}) \geq 0, \ |\text{Im}(a_{st})|^2 \leq \text{Re}(a_{st})^2 \}. \]

It is obvious that the QNQMF model (3.1) is different from the above model.

Remark 3.2 The advantages of the QNQMF model (3.1) are as follows:

1. when \( l \ll \frac{mn}{m+n} \), the storage space of \( W \) and \( H \) is smaller than the storage space of \( X \);
2. when \( X \), \( W \) and \( H \) are pure quaternion matrices, this means that the RGB image can be decomposed into two small scale RGB image. However, this factorization is sharp. Hence, it will provide room for \( W \) and \( H \) when the real component does not require the non-negative constraint;
3. it takes full advantage of color information between channels for color images when pure quaternion matrices are used to represent color images.

3.1 Quaternion Optimization Problem

To solve QNQMF (3.1), we consider the following quaternion optimization problem:

\[
\begin{align*}
\min_{W, H} \ f(W, H) &= \frac{1}{2} \| X - WH \|^2_F, \\
\text{s.t.} \ W &\in \mathbb{Q}_S^{m \times l}, \ H \in \mathbb{Q}_S^{l \times n}.
\end{align*}
\]

(3.2)

Clearly, the solution of (3.2) is exactly the solution of (3.1) if the objection function achieves zero. So the core work of solving the QNQMF becomes to handle the quaternion optimization problem (3.2), which provides room for the development of new optimization methods.

The object function \( f(W, H) \) is a real-valued function of two quaternion variables. According to [6], the gradient with respect to each quaternion variable is defined by

\[
\begin{align*}
\nabla_W f(W, H) &= \frac{\partial f(W, H)}{\partial W_0} + \frac{\partial f(W, H)}{\partial W_1} i + \frac{\partial f(W, H)}{\partial W_2} j + \frac{\partial f(W, H)}{\partial W_3} k, \\
\nabla_H f(W, H) &= \frac{\partial f(W, H)}{\partial H_0} + \frac{\partial f(W, H)}{\partial H_1} i + \frac{\partial f(W, H)}{\partial H_2} j + \frac{\partial f(W, H)}{\partial H_3} k.
\end{align*}
\]

(3.3)

(3.4)

From the Karush-Kuhn-Tucker (KKT) optimality conditions, it is not difficult to prove that \((\hat{W}, \hat{H})\) is a stationary point of (3.2) if and only if

\[
\begin{align*}
\text{Re} \ \nabla_W f(\hat{W}, \hat{H}) &= 0, \quad \hat{W} \in \mathbb{Q}_S^{m \times l}, \quad \nabla_W f(\hat{W}, \hat{H}) \in \mathbb{Q}_S^{m \times l}, \\
\text{Re} \ \nabla_H f(\hat{W}, \hat{H}) &= 0, \quad \hat{H} \in \mathbb{Q}_S^{l \times n}, \quad \nabla_H f(\hat{W}, \hat{H}) \in \mathbb{Q}_S^{l \times n}, \\
\text{Im}_i \ \hat{W} \circ \text{Im}_i \ \nabla_W f(\hat{W}, \hat{H}) &= 0, \quad \text{Im}_i \ \hat{H} \circ \text{Im}_i \ \nabla_H f(\hat{W}, \hat{H}) = 0, \\
\text{Im}_j \ \hat{W} \circ \text{Im}_j \ \nabla_W f(\hat{W}, \hat{H}) &= 0, \quad \text{Im}_j \ \hat{H} \circ \text{Im}_j \ \nabla_H f(\hat{W}, \hat{H}) = 0, \\
\text{Im}_k \ \hat{W} \circ \text{Im}_k \ \nabla_W f(\hat{W}, \hat{H}) &= 0, \quad \text{Im}_k \ \hat{H} \circ \text{Im}_k \ \nabla_H f(\hat{W}, \hat{H}) = 0.
\end{align*}
\]

(3.5)

where \( \text{Im}_i, \text{Im}_j \) and \( \text{Im}_k \) represent three imaginary parts of quaternion matrix. The proof of this assertion will be given in “Appendix A’’.

The above analysis inspires us to develop the quaternion projected gradient algorithm to handle (3.2). Before this, we need to provide four important lemmas, which will be used to prove the convergence of these algorithms.

From the definitions in (2.1), (3.3) and (3.3), the expressions of the gradient with respect to \( W \) and \( H \) are derived for the object function in the quaternion optimization problem (3.2).
Lemma 3.2 Let the real-valued function $f$ be defined as in (3.2). Then we have

$$
\begin{align*}
\nabla_W f(W, H) &= -(X - WH)H^*, \\
\nabla_H f(W, H) &= -W^*(X - WH).
\end{align*}
$$

In the KKT optimality conditions, the equations in (3.5) can be further simplified.

Lemma 3.3 The point $(\widehat{W}, \widehat{H}) \in \mathbb{Q}_+^{m \times l} \times \mathbb{Q}_+^{l \times n}$ is the stationary point of (3.2) if and only if

$$
\begin{align*}
\text{Re} \langle \nabla_W f(\widehat{W}, \widehat{H}), Y - \widehat{W} \rangle &\geq 0, \quad \forall \ Y \in \mathbb{Q}_+^{m \times l}, \\
\text{Re} \langle \nabla_H f(\widehat{W}, \widehat{H}), Z - \widehat{H} \rangle &\geq 0, \quad \forall \ Z \in \mathbb{Q}_+^{l \times n}.
\end{align*}
$$

3.2 Quaternion Projected Gradient Algorithm

To solve the optimization problem (3.2), we can optimize the object function with respect to one variable with fixing other quaternion variables:

$$
\begin{align*}
W_{r+1} &\leftarrow \arg\min_{W} \|X - WH_r\|_F^2, \\
H_{r+1} &\leftarrow \arg\min_{H} \|X - WH_{r+1}\|_F^2,
\end{align*}
$$

where $W_r$ and $H_r$ denote the solutions at iterate $r$.

Now, we consider a new quaternion projected gradient algorithm, which takes advantage of the projected gradient algorithm and the simple and effective Armijo line search for the step size [2].
Algorithm 1 (Quaternion projected gradient algorithm for QNQM (3.1))

Step 0. Given $X \in \mathbb{Q}_q^{m \times n}$, $0 < \rho < 1$ and $0 < \sigma < 1$. Initialize any feasible $W_0 \in \mathbb{Q}_q^{m \times l}$ and $H_0 \in \mathbb{Q}_q^{l \times n}$. Set $r := 0$.

Step 1. Compute
$$\nabla_W f(W_r, H_r) = -(X - W_rH_r)H_r^\ast.$$

Update
$$W_{r+1} = \mathcal{P}_{\mathbb{Q}_q^{m \times l}}[W_r - \alpha_r \nabla_W f(W_r, H_r)],$$
where $\alpha_r = \rho^s r$ and $s_r$ is the smallest non-negative integer $s$ for which
$$f(W_{r+1}, H_r) - f(W_r, H_r) \leq \sigma \text{Re} [\langle \nabla_W f(W_r, H_r), W_{r+1} - W_r \rangle].$$

(3.8)

Step 2. Compute
$$\nabla_H f(W_r, H_r) = -W_{r+1}^\ast (X - W_{r+1}H_r).$$

Update
$$H_{r+1} = \mathcal{P}_{\mathbb{Q}_q^{l \times n}}[H_r - \beta_r \nabla_H f(W_{r+1}, H_r)],$$
where $\beta_r = \rho^t r$ and $t_r$ is the smallest non-negative integer $t$ for which
$$f(W_{r+1}, H_{r+1}) - f(W_r, H_r) \leq \sigma \text{Re} [\langle \nabla_H f(W_{r+1}, H_r), H_{r+1} - H_r \rangle].$$

(3.9)

Step 3. If the stop termination criteria is satisfied, break; else, set $r := r + 1$, go to Step 1.

For Algorithm 1, we can obtain the following results.

Lemma 3.5 In Algorithm 1, we have

$$\text{Re} [\langle \nabla_W f(W_r, H_r), W_{r+1} - W_r \rangle] \leq 0,$$

(3.10)

$$\text{Re} [\langle \nabla_H f(W_{r+1}, H_r), H_{r+1} - H_r \rangle] \leq 0.$$  

(3.11)

Proof If $\alpha_r = 0$, then $W_{r+1} = W_r$ and (3.10) holds. If $\alpha_r > 0$, from Lemma 3.4 (3), we have

$$\text{Re} [\langle \mathcal{P}_{\mathbb{Q}_q^{m \times l}}[W_r - \alpha_r \nabla_W f(W_r, H_r)] - \mathcal{P}_{\mathbb{Q}_q^{m \times l}}(W_r), W_r - \alpha_r \nabla_W f(W_r, H_r) - W_r \rangle]$$

$$= \text{Re} [\langle W_{r+1} - W_r, -\alpha_r \nabla_W f(W_r, H_r) \rangle] \geq 0.$$ 

Hence, it follows that $\text{Re} [\langle \nabla_W f(W_r, H_r), W_{r+1} - W_r \rangle] \leq 0$.

Similarly, (3.11) is also tenable. This completes the proof. \(\square\)

Lemma 3.6 For given $W, D_W \in \mathbb{Q}_q^{m \times l}$ and $H, D_H \in \mathbb{Q}_q^{l \times n}$, the functions defined by

$$\theta(\alpha) := \frac{\|\mathcal{P}_{\mathbb{Q}_q^{m \times l}}(W - \alpha D_W) - W\|_F}{\alpha}, \quad \alpha > 0$$

and

$$\vartheta(\beta) := \frac{\|\mathcal{P}_{\mathbb{Q}_q^{l \times n}}(H - \beta D_H) - H\|_F}{\beta}, \quad \beta > 0$$

are nonincreasing.
Proof We firstly prove the following fact that if \( \Re \{ \langle \mathbf{V}, \mathbf{U} - \mathbf{V} \rangle \} > 0 \), then
\[
\frac{\| \mathbf{U} \|_F}{\| \mathbf{V} \|_F} \leq \frac{\Re \{ \langle \mathbf{U}, \mathbf{U} - \mathbf{V} \rangle \}}{\Re \{ \langle \mathbf{V}, \mathbf{U} - \mathbf{V} \rangle \}}
\] (3.12)
for any \( \mathbf{U}, \mathbf{V} \in \mathbb{C}^{m \times l} \). In fact, let
\[
\mathbf{U} = U_0 + U_1 \mathbf{i} + U_2 \mathbf{j} + U_3 \mathbf{k}, \quad \mathbf{V} = V_0 + V_1 \mathbf{i} + V_2 \mathbf{j} + V_3 \mathbf{k}
\]
and
\[
\mathbf{u} = \begin{pmatrix} \text{vec}(U_0) \\ \text{vec}(U_1) \\ \text{vec}(U_2) \\ \text{vec}(U_3) \end{pmatrix} \in \mathbb{R}^{4ml}, \quad \mathbf{v} = \begin{pmatrix} \text{vec}(V_0) \\ \text{vec}(V_1) \\ \text{vec}(V_2) \\ \text{vec}(V_3) \end{pmatrix} \in \mathbb{R}^{4ml},
\]
where \( \text{vec}(\cdot) \) is the straighten operator by column. Note that
\[
\Re \{ \langle \mathbf{V}, \mathbf{U} \rangle \} = \Re \{ \langle \mathbf{U}, \mathbf{V} \rangle \} = v^T u, \quad \| \mathbf{U} \|_F^2 = \| u \|_2^2, \quad \| \mathbf{V} \|_F^2 = \| v \|_2^2.
\]
Based on the Cauchy-Schwarz inequality \( u^T v \leq \| u \|_2 \| v \|_2 \), it follows that
\[
\Re \{ \langle \mathbf{V}, \mathbf{U} \rangle \} \leq \| \mathbf{U} \|_F \| \mathbf{V} \|_F, \quad \forall \mathbf{U}, \mathbf{V} \in \mathbb{C}^{m \times l}.
\]
If \( \Re \{ \langle \mathbf{V}, \mathbf{U} - \mathbf{V} \rangle \} > 0 \), we have
\[
\frac{\| \mathbf{U} \|_F}{\| \mathbf{V} \|_F} \leq \frac{\Re \{ \langle \mathbf{U}, \mathbf{U} - \mathbf{V} \rangle \}}{\Re \{ \langle \mathbf{V}, \mathbf{U} - \mathbf{V} \rangle \}}
\]
\[
\Leftrightarrow \| \mathbf{U} \|_F \Re \{ \langle \mathbf{V}, \mathbf{U} - \mathbf{V} \rangle \} \leq \| \mathbf{V} \|_F \Re \{ \langle \mathbf{U}, \mathbf{U} - \mathbf{V} \rangle \}
\]
\[
\Leftrightarrow \| \mathbf{U} \|_F \Re \{ \langle \mathbf{V}, \mathbf{U} \rangle \} + \| \mathbf{V} \|_F \Re \{ \langle \mathbf{U}, \mathbf{V} \rangle \} \leq \| \mathbf{U} \|_F \| \mathbf{V} \|_F (\| \mathbf{U} \|_F + \| \mathbf{V} \|_F)
\]
\[
\Leftrightarrow \Re \{ \langle \mathbf{V}, \mathbf{U} \rangle \} \leq \| \mathbf{U} \|_F \| \mathbf{V} \|_F.
\]
Next, we prove that \( \theta(\alpha) \) is a nonincreasing function for \( \alpha > 0 \). Let \( \alpha_1 > \alpha_2 > 0 \) be given.

Case 1: if \( \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) = \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) \), then it has \( \theta(\alpha_1) \leq \theta(\alpha_2) \).

Case 2: if \( \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) \neq \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) \), then let
\[
\mathbf{U} = \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) - \mathbf{W}, \quad \mathbf{V} = \mathcal{P}_{Q_{m \times l}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) - \mathbf{W}.
\]
From Lemma 3.4 (1), we have
\[
\Re \{ \langle \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) - (\mathbf{W} - \alpha_1 \mathbf{D}_W), \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) - \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) \rangle \} \geq 0,
\]
that is
\[
\Re \{ \langle \mathbf{U}, \mathbf{V} \rangle \} + \alpha_1 \Re \{ \langle \mathbf{D}_W, \mathbf{U} \rangle \} \geq 0
\]
\[
\Leftrightarrow \Re \{ \langle \mathbf{U}, \mathbf{V} \rangle \} \leq \alpha_1 \Re \{ \langle \mathbf{D}_W, \mathbf{U} \rangle \}.
\] (3.13)
Again, we have
\[
\Re \{ \langle \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) - (\mathbf{W} - \alpha_2 \mathbf{D}_W), \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_1 \mathbf{D}_W) - \mathcal{P}_{Q_{m \times n}}(\mathbf{W} - \alpha_2 \mathbf{D}_W) \rangle \} \geq 0,
\]
that is
\[
\text{Re} \left[ \langle V, U - V \rangle \right] + \alpha_2 \text{Re} \left[ \langle D_W, U - V \rangle \right] \geq 0
\]
\[
\Leftrightarrow \text{Re} \left[ \langle V, U - V \rangle \right] \geq \alpha_2 \text{Re} \left[ \langle D_W, V - U \rangle \right].
\] (3.14)

Since \( \alpha_1 > \alpha_2 > 0 \) and \( P_{\eta}^{m \times l}(W - \alpha_1 D_W) \neq P_{\eta}^{m \times l}(W - \alpha_2 D_W) \), thus
\[
\text{Re} \left[ P_{\eta}^{m \times l}(W - \alpha_2 D_W) - P_{\eta}^{m \times l}(W - \alpha_1 D_W), (\alpha_1 - \alpha_2)D_W \right] > 0,
\]

it means that
\[
\text{Re} \left[ \langle D_W, V - U \rangle \right] > 0.
\] (3.15)

From (3.12), (3.13), (3.14) and (3.15), it has
\[
\frac{\| P_{\eta}^{m \times l}(W - \alpha_1 D_W) - W \|_F}{\| P_{\eta}^{m \times l}(W - \alpha_2 D_W) - W \|_F} \leq \frac{\text{Re} \left[ \langle U, U - V \rangle \right]}{\text{Re} \left[ \langle V, U - V \rangle \right]} \leq \frac{\alpha_1}{\alpha_2},
\]

then it has \( \theta(\alpha_1) \leq \theta(\alpha_2) \).

Similarly, \( \theta(\beta) \) is a nonincreasing function for \( \beta > 0 \). This completes the proof. \( \Box \)

For simplicity of analysis, denote
\[
W_r(\alpha) := P_{\eta}^{m \times l}[W_r - \alpha \nabla_W f(W_r, H_r)],
\]
\[
H_r(\beta) := P_{\eta}^{n \times n}[H_r - \beta \nabla_H f(W_{r+1}, H_r)].
\]

**Lemma 3.7** Algorithm 1 is well defined.

**Proof** We only prove that there exist \( \alpha_W \) and \( \beta_H \) such that
\[
\text{Re} \left[ \langle W(\alpha) - W, W - \alpha \nabla_W f(W, H) - W(\alpha) \rangle \right] \geq 0.
\]

Hence, for all \( \alpha \in (0, 1) \), it follows Lemma 3.6 that
\[
\text{Re} \left[ \langle \nabla_W f(W, H), W - W(\alpha) \rangle \right] \geq \frac{\| W(\alpha) - W \|_F^2}{\alpha} \geq \| W(1) - W \|_F \| W(\alpha) - W \|_F.
\] (3.18)

By the mean value theorem, we have
\[
f(W(\alpha), H) - f(W, H) = \text{Re} \left[ \langle \nabla_W f(W(\xi_\alpha), H), W(\alpha) - W \rangle \right]
\]
\[
= \text{Re} \left[ \langle \nabla_W f(W, H), W(\alpha) - W \rangle \right]
\]
\[
+ \text{Re} \left[ \langle \nabla_W f(W(\xi_\alpha), H) - \nabla_W f(W, H), W(\alpha) - W \rangle \right].
\] (3.19)
Lemma 3.8 The sequence \( \{ f(W_r, H_r) \}_{r=0}^{\infty} \) generated by Algorithm 1 is monotonically non-increasing.

**Proof** From Lemma 3.5, we have

\[
f(W_{r+1}, H_r) - f(W_r, H_r) \leq 0, \quad f(W_{r+1}, H_{r+1}) - f(W_{r+1}, H_r) \leq 0.
\]

Hence, it can get that

\[
f(W_{r+1}, H_{r+1}) - f(W_r, H_r) \leq 0.
\]

This completes the proof. \( \square \)

Theorem 3.9 Let \( \{ (W_r, H_r) \}_{r=0}^{\infty} \) be the sequence generated by Algorithm 1. If

\[
\lim_{r \to +\infty} W_r = \hat{W}, \quad \lim_{r \to +\infty} H_r = \hat{H},
\]

then \( (\hat{W}, \hat{H}) \) is a stationary point of QNQMF (3.1).

**Proof** Since \( \{ f(W_r, H_r) \}_{r=0}^{\infty} \) is monotonically non-increasing, then it holds that \( f(W_r, H_r) \to f(\hat{W}, \hat{H}) \) as \( r \to +\infty \). We prove this theorem from two cases.

Case 1: \( \lim \inf_{r \to +\infty} \alpha_r \geq \hat{\alpha} > 0 \). From (3.8) and (3.18), for the sufficiently large \( r \), we have

\[
f(W_r, H_r) - f(W_{r+1}, H_r) \geq \sigma \Re \{ (\nabla_W f(W_r, H_r), W_r - W_{r+1}) \}
\geq \frac{\sigma \alpha_r \| W_{r+1} - W_r \|^2_F}{\alpha_r^2} \geq \sigma \alpha_r \| P_{Q^m} \nabla W f(W_r, H_r) - W_r \|^2_F.
\]

Taking the limit as \( r \to +\infty \), we can obtain that

\[
0 \geq \sigma \hat{\alpha} \| P_{Q^m} \nabla W f(\hat{W}, \hat{H}) - \hat{W} \|^2_F.
\]

Hence, it has \( \hat{W} = P_{Q^m} \nabla W f(\hat{W}, \hat{H}) \).

Case 2: \( \lim \inf_{r \to +\infty} \alpha_r = 0 \). Then there exists a subsequence \( \{ \alpha_r \}_{r \in \bar{R}} \) with the set \( \bar{R} \subseteq \{ 0, 1, 2, \ldots \} \) converging to zero. For all \( r \in \bar{R} \), which are sufficiently large, (3.8) will be failed at least once and therefore

\[
f(W_r(\rho^{-1} \alpha_r), H_r) - f(W_r, H_r) > \sigma \Re \{ (\nabla_W f(W_r, H_r), W_r(\rho^{-1} \alpha_r) - W_r) \}.
\]

(3.21)

Hence, \( W_r \neq W_r(\rho^{-1} \alpha_r) \); otherwise, it will contradict with (3.21). Thus, it holds \( \| W_r - W_r(\rho^{-1} \alpha_r) \|_F > 0 \).
By the mean value theorem, we have
\[
f(W_r(\rho^{-1} \alpha_r), H_r) - f(W_r, H_r) \\
= \text{Re}[(\nabla f(W_r(\xi_{\alpha_r}), H_r), W_r(\rho^{-1} \alpha) - W_r)] \\
= \text{Re}[(\nabla f(W_r, H_r), W_r(\rho^{-1} \alpha_r) - W_r)] \\
+ \text{Re}[(\nabla f(W_r(\xi_{\alpha_r}), H_r) - \nabla f(W_r, H_r), W_r(\rho^{-1} \alpha_r) - W_r)],
\]
where $W_r(\xi_{\alpha_r})$ lies in the line segment joining $W_r$ and $W_r(\rho^{-1} \alpha_r)$. Combine (3.21) and (3.22), it has
\[
(1 - \sigma) \text{Re}[(\nabla f(W_r, H_r), W_r - W_r(\rho^{-1} \alpha_r))] \\
< \text{Re}[(\nabla f(W_r(\xi_{\alpha_r}), H_r) - \nabla f(W_r, H_r), W_r(\rho^{-1} \alpha_r) - W_r)].
\]
According to Lemma 3.4 (1) and Lemma 3.6, it follows that
\[
\text{Re}[(\nabla f(W_r, H_r), W_r - W_r(\rho^{-1} \alpha_r))] \\
\geq \frac{||W_r(\rho^{-1} \alpha_r) - W_r||^2_F}{\rho^{-1} \alpha_r} \\
\geq ||W_r(1) - W_r||_F ||W_r(\rho^{-1} \alpha_r) - W_r||_F.
\]
Hence, it has
\[
(1 - \sigma)||W_r(1) - W_r||_F ||W_r(\rho^{-1} \alpha_r) - W_r||_F \\
< \text{Re}[(\nabla f(W_r(\xi_{\alpha_r}), H_r) - \nabla f(W_r, H_r), W_r(\rho^{-1} \alpha_r) - W_r)] \\
\leq ||\nabla f(W_r(\xi_{\alpha_r}), H_r) - \nabla f(W_r, H_r)||_F ||W_r(\rho^{-1} \alpha_r) - W_r||_F.
\]
We obtain that
\[
(1 - \sigma)||W_r(1) - W_r||_F < ||\nabla f(W_r(\xi_{\alpha_r}), H_r) - \nabla f(W_r, H_r)||_F. \quad (3.23)
\]
Since $\alpha_r \to 0$ and $W_r \to \hat{W}$ as $r \to +\infty$ and $r \in \hat{R}$, it follows that $W_r(\xi_{\alpha_r}) \to \hat{W}$. Taking the limit in (3.23) as $r \to +\infty$, we can obtain $(1 - \sigma)||\hat{W}(1) - \hat{W}||_F \leq 0$ that is
\[
\hat{W} = P_{Q^m \times 1}(\hat{W} - \nabla f(\hat{W}, \hat{H})).
\]
From Lemma 3.4 (1) and $\hat{W} = P_{Q^m \times 1}(\hat{W} - \nabla f(\hat{W}, \hat{H}))$, for any $Y \in Q^m$ \times 1, it follows that
\[
\text{Re}[(\nabla f(\hat{W}, \hat{H}), Y - \hat{W})] \\
= \text{Re}[(\nabla f(\hat{W}, \hat{H}), Y - \hat{W})] \\
= \text{Re}[(\nabla f(\hat{W}, \hat{H}), Y - \hat{W})] \geq 0.
\]
Similarly, it also has
\[
\text{Re}[(\nabla f(\hat{W}, \hat{H}), Z - \hat{H})] \geq 0, \quad \forall Z \in Q^l \times n.
\]
According to Lemma 3.3, it gets that $(\hat{W}, \hat{H})$ is the stationary point. This completes the proof.

Algorithm 1 is well defined and produces a monotonic non-increasing sequence. In fact, Algorithm 1 may cost the most time to search the step sizes $\alpha_r$ and $\beta_r$. So one should check as few step sizes as possible. In practice, $\alpha_r$ and $\beta_r$ may be similar to the previous step sizes $\alpha_{r-1}$ and $\beta_{r-1}$, it can use $\alpha_{r-1}$ and $\beta_{r-1}$ as the initial guesses and then either increases or decreases them in order to find the largest $\alpha_r$ and $\beta_r$ satisfying (3.8) and (3.9), respectively. Sometimes, a larger step more effectively projects variables to bounds at one
Algorithm 2 (QIPG algorithm for QNQMF (3.1))

Step 0. Given \( X \in \mathbb{Q}_{m \times n}^\dagger \), \( 0 < \rho < 1 \) and \( 0 < \sigma < 1 \). Initialize any feasible \( W_0 \in \mathbb{Q}_{m \times l}^\dagger \) and \( H_0 \in \mathbb{Q}_{l \times n}^\dagger \). Set \( r := 0 \), \( \alpha_{-1} = 0 \) and \( \beta_{-1} = 0 \).

Step 1. (1) Compute

\[
\nabla_W f(W_r, H_r) = -(X - W_r H_r) H_r^\ast.
\]

(2) Set \( \alpha_r \leftarrow \alpha_{r-1} \). If \( \alpha_r \) satisfies (3.8), repeatedly increase it by

\[
\alpha_r \leftarrow \alpha_r / \rho
\]

until \( \alpha_r \) does not satisfy (3.8). Else, repeatedly decrease\( \alpha_r \) by

\[
\alpha_r \leftarrow \alpha_r \cdot \rho
\]

until \( \alpha_r \) satisfies (3.8).

(3) Update

\[
W_{r+1} = \mathcal{P}_{\mathbb{Q}_{m \times l}^\dagger}[W_r - \alpha_r \nabla_W f(W_r, H_r)].
\]

Step 2. (1) Compute

\[
\nabla_H f(W_{r+1}, H_r) = -W_{r+1}^\ast(X - W_{r+1} H_r).
\]

(2) Set \( \beta_r \leftarrow \beta_{r-1} \). If \( \beta_r \) satisfies (3.9), repeatedly increase it by

\[
\beta_r \leftarrow \beta_r / \rho
\]

until \( \beta_r \) does not satisfy (3.9). Else, repeatedly decrease \( \beta_r \) by

\[
\beta_r \leftarrow \beta_r \cdot \rho
\]

until \( \beta_r \) satisfies (3.9).

(3) Update

\[
H_{r+1} = \mathcal{P}_{\mathbb{Q}_{l \times n}^\dagger}[H_r - \beta_r \nabla_H f(W_{r+1}, H_r)].
\]

Step 3. If the stop termination criteria is satisfied, break; else, set \( r := r + 1 \), go to Step 1.

Remark 3.3 Algorithm 1 is a general version of the projected gradient methods for the non-negative matrix factorization (2.2) proposed by Lin in [26]. The similar theoretical results can be obtained via that in [4, 9].

Remark 3.4 From Algorithms 1 and 2, it can see that the non-negativity of the imaginary parts of \( W \) and \( H \) that representing color channels are maintained during the iteration.
4 Quaternion Alternating Direction Method of Multipliers

In this section, we present a novel equivalent model of QNQMF and then solve it based on the idea of the alternating direction method of multipliers (ADMM).

The ADMM framework was originally proposed by Glowinski and Marrocco in [10]. Recently, it becomes increasingly popular in machine learning and signal processing. In particular, Boyd et al. [3] firstly proposed the ADMM to solve the non-convex NMF problem.

In order to utilize the ADMM framework, we consider the following equivalent programming problem of QNQMF (3.2)

\[
\begin{align*}
\min_{W, H} f(W, H) &= \frac{1}{2} \|X - WH\|_F^2, \\
\text{s.t. } W &= U, \quad H = V, \quad U \in \mathbb{Q}^{m \times l}, \quad V \in \mathbb{Q}^{l \times n}. 
\end{align*}
\]  

(4.1)

Then, the augmented Lagrangian function of (4.1) is

\[
\mathcal{L}_A(W, H, U, V, \Lambda, \Pi) = \frac{1}{2} \|X - WH\|_F^2 \\
- \text{Re} \left[ \langle \Lambda, W - U \rangle \right] - \text{Re} \left[ \langle \Pi, H - V \rangle \right] + \frac{\alpha}{2} \|W - U\|_F^2 + \frac{\beta}{2} \|H - V\|_F^2 
\]

(4.2)

with \( U \in \mathbb{Q}^{m \times l} \) and \( V \in \mathbb{Q}^{l \times n} \), where \( \Lambda \in \mathbb{Q}^{m \times l} \) and \( \Pi \in \mathbb{Q}^{l \times n} \) are Lagrangian multipliers and \( \alpha, \beta > 0 \) are penalty parameters.

ADMM for (4.1) is given as follows

\[
\begin{align*}
W_{r+1} &\leftarrow \arg\min_{W \in \mathbb{Q}^{m \times l}} \mathcal{L}_A(W, H_r, U_r, V_r, \Lambda_r, \Pi_r), \\
H_{r+1} &\leftarrow \arg\min_{H \in \mathbb{Q}^{l \times n}} \mathcal{L}_A(W_{r+1}, H, U_r, V_r, \Lambda_r, \Pi_r), \\
U_{r+1} &\leftarrow \arg\min_{U \in \mathbb{Q}^{m \times l}} \mathcal{L}_A(W_{r+1}, H_{r+1}, U, V_r, \Lambda_r, \Pi_r), \\
V_{r+1} &\leftarrow \arg\min_{V \in \mathbb{Q}^{l \times n}} \mathcal{L}_A(W_{r+1}, H_{r+1}, U_{r+1}, V, \Lambda_r, \Pi_r), \\
\Lambda_{r+1} &\leftarrow \Lambda_r - \alpha(W_{r+1} - U_{r+1}), \\
\Pi_{r+1} &\leftarrow \Pi_r - \beta(H_{r+1} - V_{r+1}). 
\end{align*}
\]

(4.3)

Note that the gradients with respect to each quaternion variable of the augmented Lagrangian function \( \mathcal{L}_A \) are

\[
\begin{align*}
\nabla_W \mathcal{L}_A &= (X - WH)(-H^*) - \Lambda + \alpha(W - U) \\
&= W(\mathbf{H}\mathbf{H}^* + \alpha I) - (\mathbf{X}\mathbf{H}^* + \Lambda + \alpha U); \\
\nabla_H \mathcal{L}_A &= (-W^*)(X - WH) - \Pi + \beta(H - V) \\
&= (W^*W + \beta I)H - (W^*X + \Pi + \beta V); \\
\nabla_U \mathcal{L}_A &= \Lambda - \alpha(W - U) = \alpha U - (\alpha W - \Lambda); \\
\nabla_V \mathcal{L}_A &= \Pi - \beta(H - V) = \beta V - (\beta H - \Pi). 
\end{align*}
\]

To efficiently implement (4.3), according to the first-order optimality analysis for the quaternion matrix optimization problems [32], we outline the procedure for solving those subproblems (4.3), called the quaternion alternating direction method of multipliers, denoted by QADMM.
Algorithm 3 (QADMM for QNQMF (3.1))

Step 0. Given $\mathbf{X} \in \mathbb{Q}_+^{m \times n}$ and $\alpha, \beta > 0$. Initialize any feasible $\mathbf{W}_0 \in \mathbb{Q}_+^{m \times l}$, $\mathbf{H}_0 \in \mathbb{Q}_+^{l \times n}$, $\mathbf{U}_0 \in \mathbb{Q}_+^{m \times l}$, $\mathbf{V}_0 \in \mathbb{Q}_+^{l \times n}$, $\mathbf{A}_0 \in \mathbb{Q}_+^{m \times l}$, $\mathbf{P}_0 \in \mathbb{Q}_+^{l \times n}$. Set $r := 0$.

Step 1. Update

$$
\begin{cases}
\mathbf{W}_{r+1} = (\mathbf{X} \mathbf{H}_r^* + \mathbf{A}_r + \alpha \mathbf{U}_r)(\mathbf{H}_r^* + \alpha I)^{-1}, \\
\mathbf{H}_{r+1} = (\mathbf{W}_r^* \mathbf{W}_{r+1} + \beta I)^{-1}(\mathbf{W}_{r+1} \mathbf{X} + \mathbf{P}_r + \beta \mathbf{V}_r), \\
\mathbf{U}_{r+1} = \mathcal{P}_{\mathbb{Q}_+^{m \times l}}(\mathbf{W}_{r+1} - \frac{1}{\alpha} \mathbf{A}_r), \\
\mathbf{V}_{r+1} = \mathcal{P}_{\mathbb{Q}_+^{l \times n}}(\mathbf{H}_{r+1} - \frac{1}{\beta} \mathbf{P}_r), \\
\mathbf{A}_{r+1} = \mathbf{A}_r - \alpha(\mathbf{W}_{r+1} - \mathbf{U}_{r+1}), \\
\mathbf{P}_{r+1} = \mathbf{P}_r - \beta(\mathbf{H}_{r+1} - \mathbf{V}_{r+1}).
\end{cases}
$$  \hfill (4.4)

Step 2. If the stop termination criteria is satisfied, break; else, set $r := r + 1$, go to Step 1.

Now, we provide a preliminary convergent property of the proposed QADMM.

**Theorem 4.1** Let $\{(\mathbf{W}_r, \mathbf{H}_r, \mathbf{U}_r, \mathbf{V}_r, \mathbf{A}_r, \mathbf{P}_r)\}_{r=0}^{+\infty}$ be the sequence generated by Algorithm 3. If

$$
limit_{r \to +\infty} (\mathbf{W}_r, \mathbf{H}_r, \mathbf{U}_r, \mathbf{V}_r, \mathbf{A}_r, \mathbf{P}_r) = (\hat{\mathbf{W}}, \hat{\mathbf{H}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}, \hat{\mathbf{A}}, \hat{\mathbf{P}}),
$$

then $(\hat{\mathbf{W}}, \hat{\mathbf{H}})$ is a stationary point of QNQMF (3.2).

**Proof** Firstly, we prove that

$$
\begin{cases}
\mathbf{U}_r \in \mathbb{Q}_+^{m \times l}, \\
\mathbf{A}_r \in \mathbb{Q}_+^{m \times l}, \\
\mathbf{V}_r \in \mathbb{Q}_+^{l \times n}, \\
\mathbf{P}_r \in \mathbb{Q}_+^{l \times n},
\end{cases}
$$  \hfill (4.5)

for $r = 0, 1, 2, \ldots$. In fact, denote $\mathbf{D}_r = \mathbf{W}_{r+1} - \frac{1}{\alpha} \mathbf{A}_r = \mathbf{D}_0 + \mathbf{D}_1 \mathbf{i} + \mathbf{D}_2 \mathbf{j} + \mathbf{D}_3 \mathbf{k}$, then

$$
\mathbf{U}_{r+1} = \mathbf{D}_0 + \max(D_1, 0)\mathbf{i} + \max(D_2, 0)\mathbf{j} + \max(D_3, 0)\mathbf{k} \in \mathbb{Q}_+^{m \times l}.
$$

Note that

$$
\frac{1}{\alpha} \mathbf{A}_{r+1} = \frac{1}{\alpha} \mathbf{A}_r - \mathbf{W}_{r+1} + \mathbf{U}_{r+1} = -\mathbf{D}_r + \mathbf{U}_{r+1}
$$

$$
= -\mathbf{D}_0 - \mathbf{D}_1 \mathbf{i} - \mathbf{D}_2 \mathbf{j} - \mathbf{D}_3 \mathbf{k} + \mathbf{D}_0 + \max(D_1, 0)\mathbf{i} + \max(D_2, 0)\mathbf{j} + \max(D_3, 0)\mathbf{k}
$$

$$
= (\max(D_1, 0) - D_1)\mathbf{i} + (\max(D_2, 0) - D_2)\mathbf{j} + (\max(D_3, 0) - D_3)\mathbf{k}.
$$

Since

$$
\max(a, 0) - a = \frac{1}{2}(|a| + a) - a = \frac{1}{2}(|a| - a) \geq 0
$$

for any real number $a$, it has

$$
\max(D_1, 0) - D_1 \geq 0, \quad \max(D_2, 0) - D_2 \geq 0, \quad \max(D_3, 0) - D_3 \geq 0.
$$

Thus, $\mathbf{A}_{r+1} \in \mathbb{Q}_+^{m \times l}$ as $\alpha > 0$ and $\Re \mathbf{A}_{r+1} = 0$. Again, since

$$
\max(a, 0)[\max(a, 0) - a] = \frac{1}{4}(|a| + a)(|a| - a) = 0
$$

 Springer
for any real number $a$, it follows that

$$\text{Im}_i U_r \odot \text{Im}_i \Lambda_r = 0, \quad \text{Im}_j U_r \odot \text{Im}_j \Lambda_r = 0, \quad \text{Im}_k U_r \odot \text{Im}_k \Lambda_r = 0.$$ 

Hence, the first relations of (4.5) hold. Similarly, the second relations of (4.5) also hold.

Next, taking the limit in (4.4) as $r \to +\infty$, we can obtain

\begin{align*}
\left\{ \begin{array}{l}
\hat{W}(\hat{H}^* + \alpha I) = X\hat{H}^* + \hat{\Lambda} + \alpha \hat{U}, \\
(\hat{W}^*\hat{W} + \beta I)\hat{H} = \hat{W}^*X + \hat{\Pi} + \beta \hat{V}, \\
\hat{\Lambda} = \hat{\Lambda} - \alpha (\hat{W} - \hat{U}), \\
\hat{\Pi} = \hat{\Pi} - \beta (\hat{H} - \hat{V}).
\end{array} \right.
\end{align*}

Hence, we have

\begin{align*}
\left\{ \begin{array}{l}
\hat{U} = \hat{W}, \\
\hat{\Lambda} = -(X - \hat{W}\hat{H})\hat{H}^*, \\
\hat{V} = \hat{H}, \\
\hat{\Pi} = -\hat{W}^*(X - \hat{W}\hat{H}).
\end{array} \right.
\end{align*}

(4.6)

From (4.5) and (4.6), it can get that $(\hat{W}, \hat{H})$ satisfies (3.5), that is a stationary point of QNQMF (3.2). This completes the proof. \(\square\)

**Remark 4.1** For the NMF problem (2.2), the process (4.4) reduces to

\begin{align*}
\left\{ \begin{array}{l}
W_{r+1} = (XH_r^T + \Lambda_r + \alpha U_r)(H_rH_r^T + \alpha I)^{-1}, \\
H_{r+1} = (W_{r+1}^TW_{r+1} + \beta I)^{-1}(W_{r+1}X + \Pi_r + \beta V_r), \\
U_{r+1} = \max (W_{r+1} - \frac{1}{\alpha} \Lambda_r, 0), \\
V_{r+1} = \max (H_{r+1} - \frac{1}{\beta} \Pi_r, 0), \\
\Lambda_{r+1} = \Lambda_r - \alpha (W_{r+1} - U_{r+1}), \\
\Pi_{r+1} = \Pi_r - \beta (H_{r+1} - V_{r+1}),
\end{array} \right.
\end{align*}

(4.7)

which is the algorithm proposed in [40].

**5 Applications and Numerical Experiments**

In this section, we utilize some test problems to examine the effectiveness of the proposed algorithms. All test problems are performed under MATLAB R2020b on a personal computer with 3.50 GHz central processing unit (Gen Intel(R) Core(TM) i9-11900K), 32.00 GB memory and Windows 10 operating system.

**5.1 Color Image Reconstruction**

In this subsection, we focus QNQMF on the color image reconstruction. In order to evaluate the quality of reconstructed images, we employ the peak signal to noise ratio (PSNR) criterion. The PSNR of $m \times n$ color images $X = X_Ri + X_Gj + X_Bk$ and $Z = Z_Ri + Z_Gj + Z_Bk$ is calculated as follows:

$$\text{PSNR} = 20 \log_{10} \frac{255}{\text{MSE}} [\text{dB}], \quad \text{MSE} = \sqrt{\frac{1}{mn} \sum_{s=1}^{m} \sum_{t=1}^{n} |X_{st} - Z_{st}|^2},$$

where MSE is called the mean square error. In addition, ‘Time’ denotes the elapsed CPU time in seconds.
The choice of the initial matrices

| Method    | Initial matrices                                                                 |
|-----------|----------------------------------------------------------------------------------|
| QIPGM     | $W_0 = L_1i + L_2j + L_3k$, $H_0 = S_1i + S_2j + S_3k$                          |
| QADMM     | $W_0 = U_0 = \Lambda_0 = L_1i + L_2j + L_3k$, $H_0 = V_0 = \Pi_0 = S_1i + S_2j + S_3k$ |
| RIPGM     | Red color channel: $(W_R)_0 = L_1$, $(H_R)_0 = S_1$                              |
|           | Green color channel: $(W_G)_0 = L_2$, $(H_G)_0 = S_2$                           |
|           | Blue color channel: $(W_B)_0 = L_3$, $(H_B)_0 = S_3$                            |
| RADMM     | Red color channel: $(W_R)_0 = (U_R)_0 = (\Lambda_R)_0 = L_1$, $(H_R)_0 = (V_R)_0 = (\Pi_R)_0 = S_1$ |
|           | Green color channel: $(W_G)_0 = (U_G)_0 = (\Lambda_G)_0 = L_2$, $(H_G)_0 = (V_G)_0 = (\Pi_G)_0 = S_2$ |
|           | Blue color channel: $(W_B)_0 = (U_B)_0 = (\Lambda_B)_0 = L_3$, $(H_B)_0 = (V_B)_0 = (\Pi_B)_0 = S_3$ |

When the red, green and blue color channels of the original color image are regarded as the image non-negative matrices, denoted by $X_R$, $X_G$ and $X_B$, and then the NMF can be decomposed on the image matrices $X_R$, $X_G$ and $X_B$, that is

$$X_R = W_RH_R, \quad X_G = W_GH_G, \quad X_B = W_BH_B,$$

(5.1)

where $W_R, W_G, W_B \in \mathbb{R}^{m \times l}$ and $H_R, H_G, H_B \in \mathbb{R}^{l \times n}$ are the non-negative matrices. Let $r$ be the iterations. Denote

$$Z_r = \text{Im} (W_rH_r), \quad \text{RES}(r) = \| \text{Im} X - Z_r \|_F$$

for QNQMF and

$$Z_r = (W_R)_r(H_R)_r i + (W_G)_r(H_G)_r j + (W_B)_r(H_B)_r k, \quad \text{RES}(r) = \| \text{Im} X - Z_r \|_F$$

for NMF (5.1).

Here, we compare the following four methods.

- **QIPGM**: solve QNQMF (3.1) by Algorithm 2 with $\rho = 0.01$ and $\sigma = 0.001$;
- **QADMM**: solve QNQMF (3.1) by Algorithm 3 with $\alpha = 0.01$ and $\beta = 0.01$;
- **RIPGM**: solve NMF (5.1) by [26] with $\rho = 0.01$ and $\sigma = 0.001$;
- **RADMM**: solve NMF (5.1) by (4.7) with $\alpha = 0.01$ and $\beta = 0.01$.

The choices of these numerical parameters are based on the experiments and the same values are used for the quaternion and real algorithms. Let

$$L_1 = \text{rand}(m, l), \quad L_2 = \text{rand}(m, l), \quad L_3 = \text{rand}(m, l), \quad S_1 = \text{rand}(l, n), \quad S_2 = \text{rand}(l, n), \quad S_3 = \text{rand}(l, n).$$

The choices of the initial matrices for different algorithms are presented in Table 1.

We test four different color images, named Onion, Trailer, Letters and Digits; see Fig. 1. For these color images, all algorithms run 50 iterations and the obtained numerical results are presented on Tables 2, 3, 4, and 5, respectively. Figures 2, 3, 4 and 5 present the reconstructed color images $Z_{50}$. And the relationship between the number of iterations and $\log_{10}(\text{RES}(r))$ for the QNQMF and RGB three channels NMF are plot in Fig. 6.

From these numerical results, the following facts can be observed.

- **QIPGM** and **QADMM** cost more CPU time than **RIPGM** and **RADMM** at one iteration step. However, the PSNR values of **QIPGM** and **QADMM** are better than **RIPGM** and **RADMM**, respectively. In other words, these algorithms based on QNQMF take more
FIG. 1 Experimental color images

Table 2 Numerical results of the color image Onion (135 × 198 pixels)

| l  | Method | QIPGM | QADMM | RIPGM | RADMM |
|----|--------|-------|-------|-------|-------|
| 10 | Time(s) | 0.2013 | 0.0963 | 0.0146 | 0.0064 |
|    | PSNR   | 79.8477 | 81.4198 | 79.2830 | 80.4876 |
| 20 | Time(s) | 0.2456 | 0.1434 | 0.0160 | 0.0076 |
|    | PSNR   | 80.4340 | 83.4851 | 79.4373 | 81.6233 |
| 30 | Time(s) | 0.2953 | 0.1847 | 0.0211 | 0.0100 |
|    | PSNR   | 80.7172 | 85.0360 | 79.5825 | 83.9558 |
| 40 | Time(s) | 0.3398 | 0.2243 | 0.0214 | 0.0135 |
|    | PSNR   | 80.8676 | 86.4066 | 79.4339 | 84.7264 |

Table 3 Numerical results of the color image Trailer (683 × 1024 pixels)

| l  | Method | QIPGM | QADMM | RIPGM | RADMM |
|----|--------|-------|-------|-------|-------|
| 20 | Time(s) | 9.7480 | 3.2211 | 0.8885 | 0.1091 |
|    | PSNR   | 84.7805 | 88.5377 | 85.8208 | 88.2529 |
| 40 | Time(s) | 10.2728 | 3.6850 | 0.9614 | 0.1445 |
|    | PSNR   | 84.9515 | 89.8671 | 86.2576 | 89.0544 |
| 60 | Time(s) | 11.3930 | 4.2889 | 1.0765 | 0.1748 |
|    | PSNR   | 85.0318 | 90.7518 | 86.1012 | 89.0008 |
| 80 | Time(s) | 12.0315 | 4.7420 | 0.8466 | 0.2031 |
|    | PSNR   | 85.0520 | 91.4671 | 84.6746 | 89.2806 |

CPU time at one iteration step, but they have better numerical performances than the corresponding algorithms based on NMF.

- For QIPGM and RIPGM, the sequences \( \{\text{RES}(r)\} \) are monotonically nonincreasing from Fig. 6, which is a manifestation of Lemma 3.8. Since QIPGM and RIPGM need to search the step sizes \( \alpha_r \) and \( \beta_r \) at each iteration step, it will take a lot of CPU time. Thus, the CPU time of QIPGM is the maximum of these compared algorithms.
- The PSNR value of QADMM is the best from Tables 2, 3, 4, and 5. In addition, we can see that QADMM just need few number of iterations to reach the minimum residual value \( \text{RES}(r) \); see Fig. 6. Moreover, the recovered image via QADMM is the sharpest in the image quality; see Figs. 2, 3, 4, 5. In particular, for QADMM and RADMM, it is very obvious increase in the PSNR value as the value \( l \) increases.
Table 4  Numerical results of the color image Letters ($680 \times 680$ pixels)

| $l$  |  Method | QIPGM | QADMM | RIPGM  | RADMM |
|------|---------|-------|-------|--------|-------|
| 10   | Time(s) | 5.8279| 2.0546| 0.5159 | 0.0618|
|      | PSNR    | 81.7718| **83.6386**| 81.8125| 83.5888|
| 20   | Time(s) | 6.4522| 2.1344| 0.5311 | 0.0936|
|      | PSNR    | 82.1142| **85.5912**| 81.8661| 85.3397|
| 30   | Time(s) | 6.4122| 2.1437| 0.5181 | 0.0744|
|      | PSNR    | 82.2917| **87.0045**| 82.1053| 86.4040|
| 40   | Time(s) | 6.5475| 2.2501| 0.5641 | 0.0911|
|      | PSNR    | 82.3586| **88.1478**| 82.0681| 87.2978|

Table 5  Numerical results of the color image Digits ($800 \times 800$ pixels)

| $l$  |  Method | QIPGM | QADMM | RIPGM  | RADMM |
|------|---------|-------|-------|--------|-------|
| 10   | Time(s) | 8.5460| 2.5299| 0.8968 | 0.0930|
|      | PSNR    | 83.2662| **85.0228**| 82.9325| 84.9425|
| 20   | Time(s) | 8.7719| 2.5244| 0.8720 | 0.0989|
|      | PSNR    | 83.6544| **86.7386**| 83.2849| 86.3813|
| 30   | Time(s) | 8.9382| 2.7024| 0.8939 | 0.1104|
|      | PSNR    | 83.8515| **87.9248**| 83.4227| 87.2967|
| 40   | Time(s) | 9.2815| 2.8476| 0.9283 | 0.1337|
|      | PSNR    | 83.8593| **88.9088**| 83.5650| 87.9792|

Fig. 2  Experimental results of four different algorithms for Onion with $l = 40$

5.2 Color Face Recognition

Suppose that $Q = Q_R i + Q_G j + Q_B k \in \mathbb{Q}^{m \times n}$ represents a color face image with $Q_R, Q_G$ and $Q_B$ being respectively the red, green, and blue values of pixels. Suppose that the training set contains $\mu$ known face image, denoted by $T = \{F_1, F_2, \ldots, F_\mu\}$ and the testing set contains
ν face image to be recognized, denoted by $\mathbb{S} = \{G_1, G_2, \ldots, G_\nu\}$. Here, $F_t(t = 1, 2, \ldots, \mu)$ and $G_s(s = 1, 2, \ldots, \nu)$ are $mn \times 1$ pure quaternion vectors, which are obtained via straighten $m \times n$ color face images into the columns.

Let $X = (F_1, F_2, \ldots, F_\mu)$ and $l$ be the positive integer. Assume that the basis images are $W = (w_1, w_2, \ldots, w_l)$ and the encodings are $H = (h_1, h_2, \ldots, h_\mu)$. Since the QNQMF

$$X = WH,$$

the each face $F_t$ can be approximately reconstructed by
In Algorithm 4, we present the color face recognition based on the QNQMF.

**Algorithm 4** (Color face recognition based on QNQMF)

**Step 0.** Input the color face images in $T = \{F_1, F_2, \ldots, F_\mu\}$ and $S = \{G_1, G_2, \ldots, G_\nu\}$.

Let the training data $X\text{train} = (F_1, F_2, \ldots, F_\mu)$.

**Step 1.** Do QNQMF:

$$X\text{train} = W\text{train} H\text{train}. $$

**Step 2.** For $s = 1, 2, \ldots, \nu$, compute

$$h_s = (W_{\text{train}}^* W_{\text{train}})^{-1} (W_{\text{train}}^* G_s)$$

and

$$\theta_{t,s} = \frac{\text{Re} \langle h_s, H_{\text{train}}(:, t) \rangle}{\|h_s\|_F \|H_{\text{train}}(:, t)\|_F}, \quad t = 1, \ldots, \mu. \quad (5.2)$$

Let

$$t_{\odot} = \arg\max_{1 \leq t \leq \mu} \theta_{t,s}.$$

Such $F_{t_{\odot}}$ is the optimum matching encoding for the trained image $G_s$. 

$F_t = W h_t, \quad t = 1, 2, \ldots, \mu.$
In (5.2), $H_{\text{train}}(:, t)$ denotes the $t$-th column of the quasi non-negative quaternion matrix $H_{\text{train}}$ and the similarity measure $\theta_{t,s}$ determines the matching score between $h_s$ and $H_{\text{train}}(:, t)$, which is corresponding to $G_s$ in the testing set and $F_t$ in the training set. When we test on the gray face images, $h_s$ and $H_{\text{train}}(:, t)$ reduce to the real vectors and the value

$$
\theta_{t,s} = \frac{\langle h_s, H_{\text{train}}(:, t) \rangle}{\| h_s \|_2 \| H_{\text{train}}(:, t) \|_2}
$$

is the cosine angle between the two vectors. Hence, the maximum similarity measure means the best matching trained image. We remark that the authors in [33] divided the color image into the red, green, blue color channels and proposed the following Algorithm 5.

**Algorithm 5 (Color face recognition based on color channels NMF [33])**

**Step 0.** Input the color face images in

$$
\mathbb{T} = \{ F_1, F_2, \ldots, F_\mu \} \quad \text{and} \quad \mathbb{S} = \{ G_1, G_2, \ldots, G_\nu \},
$$

where $F_t = F_{R,t}i + F_{G,t}j + F_{B,t}k$ for $t = 1, 2, \ldots, \mu$ and $G_s = G_{R,s}i + G_{G,s}j + G_{B,s}k$ for $s = 1, 2, \ldots, \nu$. Let the training matrices

$$
\begin{align*}
X_{R,\text{train}} &= (F_{R,1}, F_{R,2}, \ldots, F_{R,\mu}), \\
X_{G,\text{train}} &= (F_{G,1}, F_{G,2}, \ldots, F_{G,\mu}), \\
X_{B,\text{train}} &= (F_{B,1}, F_{B,2}, \ldots, F_{B,\mu})
\end{align*}
$$
be the red, green, and blue values of pixels, respectively.

Step 1. Do NMF on real field:

\[
\begin{align*}
X_{R,\text{train}} &= W_{R,\text{train}} H_{R,\text{train}}, \\
X_{G,\text{train}} &= W_{G,\text{train}} H_{G,\text{train}}, \\
X_{B,\text{train}} &= W_{B,\text{train}} H_{B,\text{train}}.
\end{align*}
\]

Step 2. For \( s = 1, 2, \ldots, v \), compute

\[
\begin{align*}
h_{R,s} &= (W_{R,\text{train}}^* W_{R,\text{train}})^{-1} (W_{R,\text{train}}^* G_{R,s}), \\
h_{G,s} &= (W_{G,\text{train}}^* W_{G,\text{train}})^{-1} (W_{G,\text{train}}^* G_{G,s}), \\
h_{B,s} &= (W_{B,\text{train}}^* W_{B,\text{train}})^{-1} (W_{B,\text{train}}^* G_{B,s})
\end{align*}
\]

and

\[
\theta_{t,s} = \frac{\langle h_{R,s}, H_{R,\text{train}}(:,t) \rangle}{\| h_{R,s} \|_2 \| H_{R,\text{train}}(:,t) \|_2} + \frac{\langle h_{G,s}, H_{G,\text{train}}(:,t) \rangle}{\| h_{G,s} \|_2 \| H_{G,\text{train}}(:,t) \|_2} + \frac{\langle h_{B,s}, H_{B,\text{train}}(:,t) \rangle}{\| h_{B,s} \|_2 \| H_{B,\text{train}}(:,t) \|_2}
\]

for \( t = 1, \ldots, \mu \). Let

\[
t_{\odot} = \arg\max_{1 \leq t \leq \mu} \theta_{t,s}.
\]

Such \( F_{t_{\odot}} \) is the optimum matching encoding for the trained image \( G_s \).

In this subsection, we utilize QNQM to the color face recognition. From the results presented in Subsection 5.1, we know that the numerical performance of ADMM-like method is good. Thus, we study the ADMM-like method for the face recognition.

Here, we compare the following these algorithms.

- QADMM-color: Algorithm 4 with Algorithm 3 for the color face images;
- RADMM-color: Algorithm 5 with the iteration (4.7) for the color face images;
- RADMM-gray: the iteration (4.7) with (5.3) for the gray face images;
- QPCA-color: the quaternion principal component analysis for the color face images [21].

In addition, the paraments \( \alpha = 0.01 \) and \( \beta = 0.01 \) for QADMM-color, RADMM-color and RADMM-gray methods. For QPCA-color, the value \( l \) means the first \( l \) largest eigenvalues and corresponding eigenvectors. The accuracy rate is defined as

\[
\text{Rate} = \frac{\text{True Numbers}}{\text{Total Numbers}}.
\]

**Example 1** In this experiment, we focus on the famous CASIA 3D face database [36]. CASIA 3D face database involves 123 individuals and each individual contains 37 or 38 different color face images with 640 × 480 pixels. We focus on the first 37 images for each individual. The training set is randomly selected with the random number \( \eta = 20, 24, 28, 32 \) and the remaining as the testing set. All the color face images are cropped as the 100 × 100 pixels, then resized to 50 × 50 pixels. Figure 7 presents the faces from the first one (just the first 36 images). The pictures show not only the single variations of poses, expressions and illuminations, but also the combined variations of expressions under illumination and poses under expressions. We preserve a small amount of face information for some images after processing.

For Example 1, QADMM-color, RADMM-color, RADMM-gray algorithms run 4 iterations and the accuracy rates as well as the CPU time are presented on Tables 6, 7, 8, and 9 for different dimensions \( l \) and the random numbers \( \eta \).
Table 6  Numerical results of Example 1 ($\eta = 20$)

| Method    | $l$  | 3    | 6    | 9    | 12   | 15   |
|-----------|------|------|------|------|------|------|
| QADMM-color | Time(s) | 33.2277 | 33.6415 | 34.0056 | 34.5794 | 36.7154 |
|           | Rate  | 26.88% | 51.79% | 67.10% | 70.73% | 73.17% |
| RADMM-color | Time(s) | 5.5573  | 5.6143  | 5.7008  | 5.7749  | 5.9524  |
|           | Rate  | 18.08% | 51.89% | 60.78% | 68.48% | 71.54% |
| RADMM-gray | Time(s) | 2.8677  | 2.8651  | 2.9298  | 2.9524  | 3.0534  |
|           | Rate  | 6.70%  | 38.35% | 58.97% | 63.27% | 67.53% |
| QPCA-color | Time(s) | 54.1980 | 60.3608 | 75.4179 | 87.1528 | 90.9380 |
|           | Rate  | 37.64% | 58.54% | 62.94% | 66.04% | 67.29% |

Table 7  Numerical results of Example 1 ($\eta = 24$)

| Method    | $l$  | 3    | 6    | 9    | 12   | 15   |
|-----------|------|------|------|------|------|------|
| QADMM-color | Time(s) | 56.4495 | 59.3814 | 45.3469 | 32.6791 | 58.4169 |
|           | Rate  | 24.77% | 45.65% | 63.85% | 68.04% | 72.23% |
| RADMM-color | Time(s) | 8.9977  | 9.3838  | 5.2911  | 5.3229  | 8.9831  |
|           | Rate  | 16.76% | 45.22% | 58.79% | 68.04% | 70.04% |
| RADMM-gray | Time(s) | 3.4718  | 3.6892  | 2.6669  | 2.6659  | 3.6898  |
|           | Rate  | 6.32%  | 34.83% | 54.91% | 64.23% | 67.85% |
| QPCA-color | Time(s) | 72.2730 | 100.6967 | 117.7804 | 84.2996 | 99.1015 |
|           | Rate  | 38.34% | 57.22% | 62.85% | 64.60% | 67.92% |

Example 2  In this experiment, we focus on the famous AR face database [29]. We select randomly 100 individuals (50 males and 50 females). Each individual contains 26 different color face images with $165 \times 120$ pixels, including frontal views of with different facial expressions, lighting conditions and occlusions. The training set is randomly selected with the random number $\lambda (= 14, 16, 18, 20)$ and the remaining as the testing set. All the color
Table 8 Numerical results of Example 1 \((\eta = 28)\)

| Method      | \(l\) | 3     | 6     | 9     | 12    | 15    |
|-------------|-------|-------|-------|-------|-------|-------|
| QADMM-color | Time(s) | 52.0266 | 28.9591 | 32.5345 | 47.9965 | 30.4193 |
|             | Rate   | 22.85% | 49.59% | 63.69% | 68.74% | 72.09% |
| RADMM-color | Time(s) | 6.2049  | 3.9847  | 6.6568  | 7.5933  | 4.2583  |
|             | Rate   | 14.36% | 49.59% | 60.34% | 67.21% | 71.73% |
| RADMM-gray  | Time(s) | 3.0223  | 2.0628  | 3.3289  | 2.9618  | 2.0643  |
|             | Rate   | 4.70%   | 37.85%  | 56.10%  | 66.31%  | 68.93%  |
| QPCA-color  | Time(s) | 96.3664 | 69.5055 | 84.6165 | 134.0832 | 100.5329 |
|             | Rate   | 36.04%  | 55.47%  | 61.52%  | 63.60%  | 66.21%  |

Table 9 Numerical results of Example 1 \((\eta = 32)\)

| Method      | \(l\) | 3     | 6     | 9     | 12    | 15    |
|-------------|-------|-------|-------|-------|-------|-------|
| QADMM-color | Time(s) | 22.4030 | 22.6615 | 22.8519 | 22.9498 | 23.6767 |
|             | Rate   | 24.88% | 50.57% | 69.76% | 72.68% | 76.42% |
| RADMM-color | Time(s) | 2.7697  | 2.7544  | 2.7980  | 2.7596  | 2.9234  |
|             | Rate   | 15.45% | 44.55% | 67.80% | 72.36% | 74.63% |
| RADMM-gray  | Time(s) | 1.3816  | 1.3950  | 1.4634  | 1.4897  | 1.4620  |
|             | Rate   | 5.69%   | 32.68%  | 56.91%  | 68.29%  | 71.87%  |
| QPCA-color  | Time(s) | 67.2749 | 72.5781 | 95.2822 | 147.7504 | 104.1266 |
|             | Rate   | 39.67%  | 60.49%  | 66.83%  | 69.11%  | 70.41%  |

Fig. 8 Original color face images from AR face database

Face images are resized to \(80 \times 60\) pixels. Figure 8 presents the faces from one of the male individuals.

For Example 2, the QADMM-color, RADMM-color and RADMM-gray algorithms run 4 iterations and the accuracy rates as well as the CPU time are presented on Tables 10, 11, 12, and 13 for different dimensions \(l\) and the values \(\lambda\).

From the results of Examples 1 and 2, it indicates the following facts.

- The results of Examples 1 and 2 exhibit the recognition performance and robustness of QADMM-color algorithm, which has the better accuracy rate and less CPU time than QPCA-color algorithm for the larger value \(l\). This indicates that QADMM-color algorithm has a good numerical performance.
- QPCA-color algorithm costs the most CPU time even though we use the fast Lanczos method. This is due to that we need to compute the first \(l\) largest eigenvalues and
### Table 10  Numerical results of Example 2 ($\lambda = 14$)

| Method          | $l$ | 5  | 10 | 15  | 20  | 25  |
|-----------------|-----|----|----|-----|-----|-----|
| QADMM-color     | Time(s) | 8.3088 | 8.4811 | 8.7379 | 8.9954 | 9.0350 |
|                 | Rate | 30.58% | 42.67% | 57.00% | 64.58% | 68.08% |
| RADMM-color     | Time(s) | 2.3156 | 2.3267 | 2.3678 | 2.4297 | 2.4932 |
|                 | Rate | 27.58% | 41.83% | 53.83% | 61.25% | 66.92% |
| RADMM-gray      | Time(s) | 1.0397 | 1.0343 | 1.0205 | 1.0834 | 1.0727 |
|                 | Rate | 18.92% | 33.75% | 45.92% | 54.92% | 59.67% |
| QPCA-color      | Time(s) | 13.4900 | 18.8405 | 21.8072 | 27.3111 | 32.1785 |
|                 | Rate | 33.33% | 43.42% | 46.75% | 50.58% | 53.50% |

### Table 11  Numerical results of Example 2 ($\lambda = 16$)

| Method          | $l$ | 5  | 10 | 15  | 20  | 25  |
|-----------------|-----|----|----|-----|-----|-----|
| QADMM-color     | Time(s) | 8.9551 | 8.9055 | 9.1517 | 9.3072 | 9.6715 |
|                 | Rate | 39.20% | 58.50% | 65.60% | 70.40% | 74.40% |
| RADMM-color     | Time(s) | 1.9035 | 1.9908 | 1.9718 | 2.0098 | 2.1245 |
|                 | Rate | 34.30% | 57.10% | 60.60% | 68.20% | 72.70% |
| RADMM-gray      | Time(s) | 0.9110 | 0.9376 | 0.9309 | 0.9337 | 0.9478 |
|                 | Rate | 24.70% | 47.40% | 56.00% | 62.50% | 68.60% |
| QPCA-color      | Time(s) | 13.1535 | 19.2302 | 21.6543 | 26.7626 | 31.7443 |
|                 | Rate | 45.70% | 58.80% | 63.90% | 67.40% | 68.40% |

### Table 12  Numerical results of Example 2 ($\lambda = 18$)

| Method          | $l$ | 5  | 10 | 15  | 20  | 25  |
|-----------------|-----|----|----|-----|-----|-----|
| QADMM-color     | Time(s) | 8.0949 | 8.5623 | 8.9744 | 8.7200 | 9.6274 |
|                 | Rate | 25.87% | 43.13% | 50.50% | 58.25% | 65.50% |
| RADMM-color     | Time(s) | 1.8528 | 1.8479 | 1.9877 | 1.9465 | 2.1142 |
|                 | Rate | 21.75% | 42.63% | 49.75% | 56.25% | 62.50% |
| RADMM-gray      | Time(s) | 0.8996 | 0.9225 | 1.2338 | 0.8630 | 0.9577 |
|                 | Rate | 14.62% | 34.25% | 44.12% | 51.00% | 59.25% |
| QPCA-color      | Time(s) | 13.4892 | 18.9881 | 22.0845 | 27.9783 | 32.1396 |
|                 | Rate | 33.62% | 46.37% | 49.13% | 51.50% | 52.88% |

corresponding eigenvectors of the covariance matrix from the training set, which is a $2500 \times 2500$ quaternion matrix for Example 1 or a $4800 \times 4800$ quaternion matrix for Example 2.

- When the value $l$ is small, the accuracy rates of QADMM-color and QPCA-color are much higher than RADMM-color and RADMM-gray. This indicates that the algorithms based on the quaternion are better than encoded three color channels of color images and single channel of gray level images. In other word, color images represented the pure quaternion matrices can explore the information between color channels. In addition, with the increase of the dimension $l$, the accuracy rate of face recognition also increases.
Table 13 Numerical results of Example 2 ($\lambda = 20$)

| Method       | $l$  | 5  | 10 | 15 | 20  | 25  |
|--------------|-----|----|----|----|-----|-----|
| QADMM-color  | Time(s) | 9.0368 | 9.1079 | 9.8130 | 9.9078 | 10.2441 |
|              | Rate      | 31.00% | 51.50% | 65.67% | 81.50% | 84.17% |
| RADMM-color  | Time(s) | 1.5250 | 1.5520 | 1.5598 | 1.6164 | 1.7145 |
|              | Rate      | 30.83% | 50.17% | 62.83% | 78.00% | 82.50% |
| RADMM-gray   | Time(s) | 0.7205 | 0.7291 | 0.7306 | 0.8068 | 0.8167 |
|              | Rate      | 17.67% | 40.50% | 52.33% | 59.83% | 69.17% |
| QPCA-color   | Time(s) | 13.9676 | 18.7256 | 21.7172 | 27.5337 | 33.5137 |
|              | Rate      | 34.83% | 44.67% | 50.83% | 54.67% | 55.67% |

6 Concluding Remarks

We introduced the QNQMF model and proposed two solving numerical algorithms. Moreover, we explored the advantages of QNQMF for color image reconstruction and color face recognition. From these numerical results, it can see that QNQMF has simple form, good interpretability and small storage space. And the algorithms encoded on the quaternion perform better than the algorithms encoded on the red, green and blue channels. In particular, the QNQMF model has a better face recognition ability and robustness than real NMF model encoded separately on color channels of color images and on single channel of gray level images for the same data.

Acknowledgements The authors are grateful for the Editor-in-Chief, Associate Editor, and Reviewers for their valuable comments and insightful suggestions that helped to improve this research significantly.

Author Contributions The authors have contributed equally in this scholarly work.

Funding This research is supported by the National Natural Science Foundation of China grants 62105064, 12171210 and 12090011; the Major Projects of Universities in Jiangsu Province of China grant 21KJA110001; and the Natural Science Foundation of Fujian Province of China grant 2020J05034.

Data Availability Statement Enquiries about data availability should be directed to the authors.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The Real Representation Method

In this appendix, we analysis the real representation method.

Remark that (3.1) is equivalent to

$$\Upsilon_X = \Upsilon_W \Upsilon_H.$$  (A.1)
or

\[
\begin{align*}
X_0 &= W_0 H_0 - W_1 H_1 - W_2 H_2 - W_3 H_3, \\
X_1 &= W_0 H_1 + W_1 H_0 + W_2 H_3 - W_3 H_2, \\
X_2 &= W_0 H_2 - W_1 H_3 + W_2 H_0 + W_3 H_1, \\
X_3 &= W_0 H_3 + W_1 H_2 - W_2 H_1 + W_3 H_0
\end{align*}
\]  \hspace{1cm} (A.2)

with \( W_1, W_2, W_3 \geq 0 \) and \( H_1, H_2, H_3 \geq 0 \).

For the equivalent problem (A.1), it is quite expansive to solve as the dimension-expanding problem caused by the real counterpart method. Meanwhile, it is different from NMF (2.2); thus those numerical algorithms proposed for NMF will not directly applied to solve (A.1). And for the problem (A.2), it will be difficult to comprehend adequately the structure of quaternion quasi non-negative matrices \( W \) and \( H \) if \( W_0, W_1, W_2, W_3 \) and \( H_0, H_1, H_2, H_3 \) are as separate variables.

In fact, the optimization problem (3.2) is equivalent to

\[
\begin{align*}
\min f(W, H) &= \|X_0 - (W_0 H_0 - W_1 H_1 - W_2 H_2 - W_3 H_3)\|_F^2 \\
&+ \|X_1 - (W_0 H_1 + W_1 H_0 + W_2 H_3 - W_3 H_2)\|_F^2 \\
&+ \|X_2 - (W_0 H_2 - W_1 H_3 + W_2 H_0 + W_3 H_1)\|_F^2 \\
&+ \|X_3 - (W_0 H_3 + W_1 H_2 - W_2 H_1 + W_3 H_0)\|_F^2,
\end{align*}
\]  \hspace{1cm} (A.3)

s.t. \( W_1 \geq 0, W_2 \geq 0, W_3 \geq 0, H_1 \geq 0, H_2 \geq 0, H_3 \geq 0 \).

Now, we consider the Karush-Kuhn-Tucker (KKT) optimality conditions of the optimization problem (A.3). Let \( y = (y_1^T, y_2^T, \ldots, y_8^T)^T \in \mathbb{R}^{4l(m+n)} \) with

\[
y_1 = \text{vec}(W_0), y_2 = \text{vec}(W_1), y_3 = \text{vec}(W_2), y_4 = \text{vec}(W_3), \\
y_5 = \text{vec}(H_0), y_6 = \text{vec}(H_1), y_7 = \text{vec}(H_2), y_8 = \text{vec}(H_3),
\]

and

\[
\begin{align*}
G_{1,s,t}(y) &= (W_1)_{st}, G_{2,s,t}(y) = (W_2)_{st}, G_{3,s,t}(y) = (W_3)_{st}, \\
J_{1,t,k}(y) &= (H_1)_{tk}, J_{2,t,k}(y) = (H_2)_{tk}, J_{3,t,k}(y) = (H_3)_{tk}
\end{align*}
\]

for \( s = 1, \ldots, m; t = 1, \ldots, l \) and \( k = 1, \ldots, n \).

Suppose that \( \hat{y} \) is a local minimum of the problem (A.3). Denote

\[
\begin{align*}
\mathbb{I}_G(\hat{y}) &= \{(a, s, t) \mid G_{a,s,t}(\hat{y}) = 0\}, \quad \mathbb{I}_f(\hat{y}) = \{(b, t, k) \mid J_{b,t,k}(\hat{y}) = 0\}.
\end{align*}
\]

It is obvious that \( f, G_{a,s,t}(a = 1, 2, 3; s = 1, \ldots, m; t = 1, \ldots, l) \) and \( J_{b,t,k}(b = 1, 2, 3; t = 1, \ldots, l; k = 1, \ldots, n) \) are continuously differentiable functions over \( \mathbb{R}^{4l(m+n)} \). And it is easy to verify that the gradients of the active constraints

\[
\{\nabla G_{a,s,t}(\hat{y})\}_{(a,s,t)\in \mathbb{I}_G(\hat{y})} \cup \{\nabla J_{b,t,k}(\hat{y})\}_{(b,t,k)\in \mathbb{I}_J(\hat{y})}
\]

are linearly independent. Then, there exist multipliers \( c_{a,s,t}(a = 1, 2, 3; s = 1, \ldots, m; t = 1, \ldots, l) \) and \( d_{b,t,k}(b = 1, 2, 3; t = 1, \ldots, l; k = 1, \ldots, n) \) such that

\[
\begin{align*}
\nabla f(\hat{y}) - \sum_{a,s,t} c_{a,s,t} \nabla G_{a,s,t}(\hat{y}) - \sum_{b,t,k} d_{b,t,k} \nabla J_{b,t,k}(\hat{y}) &= 0, \\
G_{a,s,t}(\hat{y}) &\geq 0, \quad c_{a,s,t} \geq 0, \quad c_{a,s,t} G_{a,s,t}(\hat{y}) = 0, \\
J_{b,t,k}(\hat{y}) &\geq 0, \quad d_{b,t,k} \geq 0, \quad d_{b,t,k} J_{b,t,k}(\hat{y}) = 0,
\end{align*}
\]  \hspace{1cm} (A.4)

where \( a, b = 1, 2, 3; s = 1, \ldots, m; t = 1, \ldots, l \) and \( k = 1, \ldots, n \). Let the matrices \( C_a = (c_{a,s,t}) \in \mathbb{R}^{m \times l} \) with \( a = 1, 2, 3 \) and \( D_b = (d_{b,t,k}) \in \mathbb{R}^{l \times n} \) with \( b = 1, 2, 3 \). Then, if

\[ Springer \]
According to (A.5), we can obtain the KKT optimality conditions of the problem (3.2), that is (3.5).

Appendix B. Proofs of Lemmas 3.2, 3.3 and 3.4

In this appendix, we present the proofs of lemmas in Sect. 3.

Define the matrix set
\[ \mathbb{R}^{m \times 4n}_+ := \{ R = (R_0, R_1, R_2, R_3) \in \mathbb{R}^{m \times 4n} | R_0 \in \mathbb{R}^{m \times n}, R_1, R_2, R_3 \in \mathbb{R}^{m \times n}_+ \}, \]
where \( \mathbb{R}^{m \times n}_+ \) is the non-negative set on the real field. Hence, the sets \( \mathbb{Q}^{m \times n}_+ \) and \( \mathbb{R}^{m \times 4n}_+ \) are isomorphic.

It is obvious that \( \mathbb{R}^{m \times 4n}_+ \) is a nonempty closed convex set, and let \( \mathcal{P}_{\mathbb{R}^{m \times 4n}_+} \) be the projection into \( \mathbb{R}^{m \times 4n}_+ \), which is defined by
\[
\mathcal{P}_{\mathbb{R}^{m \times 4n}_+}(R) := \arg\min Y - R \|_{F}, \ \forall Y \in \mathbb{R}^{m \times 4n}_+,
\]
then we can obtain that
\[
\mathcal{P}_{\mathbb{R}^{m \times 4n}_+}(R) = (R_0, \mathcal{P}_+(R_1), \mathcal{P}_+(R_2), \mathcal{P}_+(R_3)),
\]
where \( \mathcal{P}_+(R_s) = \max(R_s, 0), s = 1, 2, 3. \)

Proof of Lemma 3.2 Let \( X = X_0 + X_1 i + X_2 j + X_3 k \), \( W = W_0 + W_1 i + W_2 j + W_3 k \) and \( H = H_0 + H_1 i + H_2 j + H_3 k \). Since
\[
WH = (W_0 H_0 - W_1 H_1 - W_2 H_2 - W_3 H_3) + (W_0 H_1 + W_1 H_0 + W_2 H_3 - W_3 H_2)i
+ (W_0 H_2 - W_1 H_1 + W_2 H_0 + W_3 H_1)j + (W_0 H_3 + W_1 H_2 - W_2 H_1 + W_3 H_0)k
:= M_0 + M_1 i + M_2 j + M_3 k
\]
and
\[
\frac{1}{2} \| X - WH \|_F^2 = \frac{1}{2} (\| X_0 - M_0 \|_F^2 + \| X_1 - M_1 \|_F^2 + \| X_2 - M_2 \|_F^2 + \| X_3 - M_3 \|_F^2),
\]
according to (3.3), we can get that
\[
\nabla_W f(W, H) = \frac{\partial f(W, H)}{\partial W_0} + \frac{\partial f(W, H)}{\partial W_1} i + \frac{\partial f(W, H)}{\partial W_2} j + \frac{\partial f(W, H)}{\partial W_3} k
\]
\[
= (X_0 - M_0)(-H_0^T) + (X_1 - M_1)(-H_1^T) + (X_2 - M_2)(-H_2^T) + (X_3 - M_3)(-H_3^T)
+ [(X_0 - M_0)(H_1^T) + (X_1 - M_1)(-H_0^T) + (X_2 - M_2)(H_3^T) + (X_3 - M_3)(-H_2^T)]i
+ [(X_0 - M_0)(H_2^T) + (X_1 - M_1)(-H_3^T) + (X_2 - M_2)(-H_0^T) + (X_3 - M_3)(H_1^T)]j
\]
\[ + [(X_0 - M_0)(H_0^T) + (X_1 - M_1)(H_1^T) + (X_2 - M_2)(-H_1^T) + (X_3 - M_3)(-H_0^T)]k = - (X - WH)H^*. \]

Similarly, we can get that \( \nabla_H f(W, H) = -W^*(X - WH) \).

**Proof of Lemma 3.3** It is obvious that if the point \((\hat{W}, \hat{H})\) satisfies (3.5), then it meets (3.6). We will show next that conversely: if (3.6) holds, then the point \((\hat{W}, \hat{H})\) also meets (3.5).

Case 1: if \( \text{Re} \ nabla_W f(\hat{W}, \hat{H}) \neq 0 \), then there exist \( s_0 \) and \( t_0 \) such that \( a_0 := (\text{Re} \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} \neq 0 \). Let

\[
Y_{st} = \begin{cases} \hat{W}_{st} - a_0, & \text{if } s = s_0, \ t = t_0, \\ \hat{W}_{st}, & \text{others.} \end{cases}
\]

As \( \hat{W} \in \mathbb{Q}^{m \times l}_\hat{\eta} \), thus we have \( Y \in \mathbb{Q}^{m \times l}_\hat{\eta} \) and

\[
\text{Re} [(\nabla_W f(\hat{W}, \hat{H}), \ Y - \hat{W})] = -a_0^2 < 0,
\]

which is contradict with the first inequality of (3.6).

Case 2: if \( \nabla_W f(\hat{W}, \hat{H}) \notin \mathbb{Q}^{m \times l}_\hat{\eta} \), there exist \( s_0 \) and \( t_0 \) such that one of the following three inequalities

\[
\begin{cases} (\text{Im}_i \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} < 0, \\
(\text{Im}_j \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} < 0, \\
(\text{Im}_k \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} < 0
\end{cases}
\]

holds. Without loss of generality, assume that \( a_1 := (\text{Im}_i \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} < 0 \). Let

\[
Y_{st} = \begin{cases} \hat{W}_{st} - a_1 i, & \text{if } s = s_0, \ t = t_0, \\ \hat{W}_{st}, & \text{others.} \end{cases}
\]

Then, \( (\text{Im}_i \ Y)_{s_0, t_0} = (\text{Im}_i \ \hat{W})_{s_0, t_0} - a_1 > 0 \). Thus, we have \( Y \in \mathbb{Q}^{m \times l}_\hat{\eta} \) and

\[
\text{Re} [(\nabla_W f(\hat{W}, \hat{H}), \ Y - \hat{W})] = -a_1^2 < 0,
\]

which is contradict with the first inequality of (3.6).

Case 3: if \( \text{Im}_i \ \hat{W} \circ \text{Im}_i \ nabla_W f(\hat{W}, \hat{H}) \neq 0 \), there exist \( s_0 \) and \( t_0 \) such that

\[
\hat{w}_1 := (\text{Im}_i \ \hat{W})_{s_0, t_0} > 0, \quad a_1 := (\text{Im}_i \ nabla_W f(\hat{W}, \hat{H}))_{s_0, t_0} > 0.
\]

Let

\[
Y_{st} = \begin{cases} 0, & \text{if } s = s_0, \ t = t_0, \\ \hat{W}_{st}, & \text{others.} \end{cases}
\]

Then, we have \( Y \in \mathbb{Q}^{m \times l}_\hat{\eta} \) and

\[
\text{Re} [(\nabla_W f(\hat{W}, \hat{H}), \ Y - \hat{W})] = -\hat{w}_1 a_1 < 0,
\]

which is contradict with the first inequality of (3.6).

If \( \text{Im}_j \ \hat{W} \circ \text{Im}_j \ nabla_W f(\hat{W}, \hat{H}) \neq 0 \) and \( \text{Im}_k \ \hat{W} \circ \text{Im}_k \ nabla_W f(\hat{W}, \hat{H}) \neq 0 \), we can also show that they are contradict with the first inequality of (3.6). Hence, if the first inequality of (3.6) holds if and only if \( \hat{W} \) satisfies (3.5).

Similarly, it can verify that the second inequality of (3.6) holds if and only if \( \hat{H} \) satisfies (3.5). \qed
Proof of Lemma 3.4 Let $Y = Y_0 + Y_1 i + Y_2 j + Y_3 k \in \mathbb{Q}^{m \times n}$ and $Z = Z_0 + Z_1 i + Z_2 j + Z_3 k \in \mathbb{Q}^{m \times n}$.

We prove the result (1). If $Z \in \mathbb{Q}^{m \times n}$, it follows that

$$\text{Re} \left[ (\mathcal{P}_{\mathbb{Q}^{m \times n}}(Y) - Y, Z - \mathcal{P}_{\mathbb{Q}^{m \times n}}(Y)) \right] = \sum_{s=1}^{3} \langle \mathcal{P}_+(Y_s) - Y_s, Z_s - \mathcal{P}_+(Y_s) \rangle \geq 0,$$

which is based on the fact $\langle \mathcal{P}_+(a) - a, b - \mathcal{P}_+(a) \rangle \geq 0$ for any $a \in \mathbb{R}$ and $b \in \mathbb{R}_+$. We prove the result (2). It follows that

$$\text{Re} \left[ (\mathcal{P}_{\mathbb{Q}^{m \times n}}(Y) - \mathcal{P}_{\mathbb{Q}^{m \times n}}(Z), Y - Z) \right]$$

$$= \langle Y_0 - Z_0, Y_0 - Z_0 \rangle + \sum_{s=1}^{3} \langle \mathcal{P}_+(Y_s) - \mathcal{P}_+(Z_s), Y_s - Z_s \rangle \geq 0,$$

which is based on the fact $\langle \mathcal{P}_+(a) - \mathcal{P}_+(b), a - b \rangle \geq 0$ for any $a, b \in \mathbb{R}$. It is obvious that $\mathcal{P}_{\mathbb{Q}^{m \times n}}(Y) \neq \mathcal{P}_{\mathbb{Q}^{m \times n}}(Z)$, then the strict inequality holds.

We prove the result (3). It follows that

$$\| \mathcal{P}_{\mathbb{Q}^{m \times n}}(Y) - \mathcal{P}_{\mathbb{Q}^{m \times n}}(Z) \|_F^2 = \| Y_0 - Z_0 \|_F^2 + \sum_{s=1}^{3} \| \mathcal{P}_+(Y_s) - \mathcal{P}_+(Z_s) \|_F^2$$

$$\leq \| Y_0 - Z_0 \|_F^2 + \sum_{s=1}^{3} \| Y_s - Z_s \|_F^2 = \| Y - Z \|_F^2,$$

which is based on the fact $\| \mathcal{P}_+(a) - \mathcal{P}_+(b) \| \leq \| a - b \|$ for any $a, b \in \mathbb{R}$. \qed

References

1. Ang, A.M.S., Gillis, N.: Accelerating nonnegative matrix factorization algorithms using extrapolation. Neural Comput. 31(2), 417–439 (2019)
2. Bertsekas, D.P.: On the Goldstein-Levitin-Polyak gradient projection method. IEEE Trans. Automat. Contr. 21(2), 174–184 (1976)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011)
4. Calamai, P.P., Moré, J.J.: Projected gradient methods for linearly constrained problems: Math. Program. 39, 93–116 (1987)
5. Chen, Y., Jia, Z.G., Peng, Y., Zhang, D.: A new structure-preserving quaternion QR decomposition method for color image blind watermarking. Signal Process. 185, 108088 (2021)
6. Chen, Y.N., Qi, L.Q., Zhang, X.Z., Xu, Y.W.: A low rank quaternion decomposition algorithm and its application in color image inpainting. arXiv:2009.12203 (2020)
7. Cichocki, A., Zdunek, R.: Regularized alternating least squares algorithms for non-negative matrix/tensor factorization. Advances in International Symposium on Neural Networks, pp. 793–802. Springer, Berlin (2007)
8. Flamant, J., Miron, S., Brie, D.: Quaternion non-negative matrix factorization: Definition, uniqueness, and algorithm. IEEE Trans. Signal Process. 68, 1870–1883 (2020)
9. Gafni, E.M., Bertsekas, D.P.: Convergence of a gradient projection method. Report LIDS-P-1201, Lab. for Info. and Dec. Systems, M.I.T. (1982)
10. Glowinski, R., Marrocco, A.: Sur l’approximation par éléments finis d’ordre un, et la résolution par penalisation-dualité d’une classe de problèmes de Dirichlet nonlinéaires. ESAIM: Mathematical Modelling and Numerical Analysis Modélisation Mathématique et Analyse Numérique 9, 41–76 (1975)
11. Gong, P.H., Zhang, C.S.: Efficient nonnegative matrix factorization via projected Newton method. Pattern Recogn. 45(9), 3557–3565 (2012)
12. Guan, N.Y., Tao, D.C., Luo, Z.G., Yuan, B.: NeNMF: an optimal gradient method for nonnegative matrix factorization. IEEE Trans. Signal Process. 60, 2882–2898 (2012)
13. Hajinezhad, D., Chang, T.H., Wang, X.F., Shi, Q.J., Hong, M.Y.: Nonnegative matrix factorization using ADMM: Algorithm and convergence analysis. In International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 4742–4746 (2016)
14. Hamilton, W.R.: Elements of Quaternions. Longmans Green, London (1866)
15. Huang, C.Y., Fang, Y.Y., Wu, T.T., Zeng, T.Y., Zeng, Y.H.: Quaternion screened Poisson equation for low-light image enhancement. IEEE Signal Process. Lett. 29, 1417–1421 (2022)
16. Huang, C.Y., Li, Z., Liu, Y.B., Wu, T.T., Zeng, T.Y.: Quaternion-based weighted nuclear norm minimization for color image restoration. Pattern Recogn. 128, 108665 (2022)
17. Jia, Z.G.: The Eigenvalue Problem of Quaternion Matrix: Structure-Preserving Algorithms and Applications. Science Press, Beijing (2019)
18. Jia, Z.G., Ng, M.K.: Structure preserving quaternion generalized minimal residual method. SIAM J. Matrix Anal. Appl. 42(2), 616–634 (2021)
19. Jia, Z.G., Jin, Q.Y., Ng, M.K., Zhao, X.L.: Non-local robust quaternion matrix completion for large-scale color image and video inpainting. IEEE Trans. Image Process. 31, 3868–3883 (2022)
20. Jia, Z.G., Ng, M.K., Song, G.J.: Robust quaternion matrix completion with applications to image inpainting. Numer. Linear Algebra Appl. 26(4), e2245 (2019)
21. Jia, Z.G., Ng, M.K., Song, G.J.: Lanczos method for large-scale quaternion singular value decomposition. Numer. Algor. 82, 699–717 (2019)
22. Jia, Z.G., Ng, M.K., Wang, W.: Color image restoration by saturation-value total variation. SIAM J. Imaging Sci. 12(2), 972–1000 (2019)
23. Kim, H., Park, H.: Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. SIAM J. Matrix Anal. Appl. 30(2), 713–730 (2008)
24. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)
25. Li, J.Z., Yu, C.Y., Gupta, B.B., Ren, X.C.: Color image watermarking scheme based on quaternion Hadamard transform and Schur decomposition. Multimed. Tools Appl. 77(4), 4545–4561 (2018)
26. Lin, C.J.: Projected gradient methods for nonnegative matrix factorization. Neural Comput. 19(10), 2756–2779 (2007)
27. Liu, Q.H., Ling, S.T., Jia, Z.G.: Randomized quaternion singular value decomposition for low-rank matrix approximation. SIAM J. Sci. Comput. 44(2), A870–A900 (2022)
28. Lu, X.Q., Wu, H., Yuan, Y.: Double constrained NMF for hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens. 52(5), 2746–2758 (2013)
29. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report 24 (1998)
30. Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126 (1994)
31. Pompili, F., Gillis, N., Absil, P.A., Glineur, F.: Two algorithms for orthogonal nonnegative matrix factorization with application to clustering. Neurocomputing 141, 15–25 (2014)
32. Qi, L.Q., Luo, Z.Y., Wang, Q.W., Zhang, X.Z.: Quaternion matrix optimization: Motivation and analysis. J. Optim. Theory Appl. 193(1), 621–648 (2022)
33. Rajapakse, M., Tan, J., Rajapakse, J.C.: Color channel encoding with NMF for face recognition. 2004 International Conference on Image Processing, IEEE (2004)
34. Rapin, J., Bobin, J., Larue, A., Starck, J.L.: NMF with sparse regularizations in transformed domains. SIAM J. Imaging Sci. 7(4), 2040–2047 (2014)
35. Song, G.J., Ding, W.Y., Ng, M.K.: Low rank pure quaternion approximation for pure quaternion matrices. SIAM J. Matrix Anal. Appl. 42(1), 58–82 (2021)
36. The CASIA 3D Face Database. http://www.cbsr.ia.ac.cn/english/3dface%20databases.asp
37. Wu, T.T., Mao, Z.H., Li, Z.Y., Zeng, Y.H., Zeng, T.Y.: Efficient color image segmentation via quaternion-based $L_1/L_2$ Regularization. J. Sci. Comput. 93(1), 1–26 (2022)
38. Xiao, X.L., Zhou, Y.C.: Two-dimensional quaternion PCA and sparse PCA. IEEE Trans. Neur. Net. Lear. 30(7), 2028–2042 (2018)
39. Xu, Y.Y., Yin, W.T.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J. Imaging Sci. 6(3), 1758–1789 (2013)
40. Zhang, S.F., Huang, D.Y., Xei, L., Chng, E.S., Li, H.Z., Dong, M.H.: Non-negative matrix factorization using stable alternating direction method of multipliers for source separation. In Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), IEEE, 222–228 (2015)
41. Zhao, M.X., Jia, Z.G., Cai, Y.F., Chen, X., Gong, D.W.: Advanced variations of two-dimensional principal component analysis for face recognition. Neurocomputing 452, 653–664 (2021)
Authors and Affiliations

Yifen Ke\textsuperscript{1,2} · Changfeng Ma\textsuperscript{1,2} · Zhigang Jia\textsuperscript{3,4} · Yajun Xie\textsuperscript{5} · Riwei Liao\textsuperscript{6}

Yifen Ke
keyifen@fjnu.edu.cn

Zhigang Jia
zhgjia@jsnu.edu.cn

Yajun Xie
xyj@fzfu.edu.cn

Riwei Liao
riweiliao2020@fjnu.edu.cn

\textsuperscript{1} School of Mathematics and Statistics & FJKLMAA, Fujian Normal University, Fuzhou 350117, People’s Republic of China

\textsuperscript{2} Center for Applied Mathematics of Fujian province (FJNU), Fuzhou 350117, People’s Republic of China

\textsuperscript{3} School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, People’s Republic of China

\textsuperscript{4} The Research Institute of Mathematical Science, Jiangsu Normal University, Xuzhou 221116, People’s Republic of China

\textsuperscript{5} School of Big Data, Fuzhou University of International Studies and Trade, Fuzhou 350202, People’s Republic of China

\textsuperscript{6} College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350117, People’s Republic of China