Convex Quaternion Optimization for Signal Processing: Theory and Applications

Shuning Sun, Qiankun Diao, Dongpo Xu, Pauline Bourigault and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—Convex optimization methods have been extensively used in the fields of communications and signal processing. However, the theory of quaternion optimization is currently not as fully developed and systematic as that of complex and real optimization. To this end, we establish an essential theory of convex quaternion optimization for signal processing based on the generalized Hamilton-real (GHR) calculus. This is achieved in a way which conforms with traditional complex and real optimization theory. For rigorous, We present five discriminant theorems for convex quaternion functions, and four discriminant criteria for strongly convex quaternion functions. Furthermore, we provide a fundamental theorem for the optimality of convex quaternion optimization problems, and demonstrate its utility through three applications in quaternion signal processing. These results provide a solid theoretical foundation for convex quaternion optimization and open avenues for further developments in signal processing applications.**

***Keywords*—Convex quaternion functions, strongly convex quaternion functions, convex quaternion optimization, quaternion signal processing.**

I. Introduction

QUATERNIONS were first introduced by William Hamilton in 1843 as an associative but non-commutative algebra over the real numbers [1]. Since then, they have become a powerful tool in many fields, including image processing [2, 3], signal processing [4–6], and machine learning [7–9]. Examples include the work by Jia et al. [3], who introduced a robust method for quaternion matrix completion, that can be used to reconstruct large-scale color images. Flamant et al. [10] demonstrated the efficiency of Quaternion Fourier Transform (QFT) in processing bivariate signals. Ogunfunmi et al. [11] presented a kernel adaptive filter for quaternion data. Moreover, Mengüç et al. [12] designed quaternion-valued second-order Volterra adaptive filters for nonlinear 3-D and 4-D signal processing. Xia et al. [13] established an estimation framework for processing quaternion-valued Gaussian data. Finally, Zhang et al. [7] discussed a new method for reducing the computation cost of quaternion signal estimation. Enshaeifar et al. [14] introduced quaternion-valued singular spectrum analysis for multichannel electroencephalogram analysis.

This work was funded in part by the National Natural Science Foundation of China (No. 62176051), in part by National Key R&D Program of China (No. 2021YFA1003400), and in part by the Fundamental Research Funds for the Central Universities of China (No. 2412020FZ024). (*Corresponding author: Dongpo Xu*)

Shuning Sun, Qiankun Diao and Dongpo Xu are with the Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, 130024, China. (e-mail: xudp100@nenu.edu.cn)

Pauline Bourigault and Danilo P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK. (e-mail: d.mandic@imperial.ac.uk)

The theory of real-valued and complex-valued convex optimization is well-established and has seen widely used in the areas of communications [15], machine learning [16–18] and signal processing [19–21]. In recent years, convex quaternion optimization has also attracted interest. For example, Qi et al. [2] studied first-order derivatives and second-order partial derivatives of real-valued functions of quaternion variables over their real and imaginary i, j, k parts. However, this complicates the proof and computational process in quaternion optimization. Flamant et al. [22] and Liu et al. [23] provided first-order characterization of quaternion functions by generalized Hamilton-real (GHR) calculus [24]. However, these useful attempts lack the discussion of gradient monotonicity and second-order characterization for convex quaternion function, a pre-requisite for practical applications.

To fill this void, we have systematically address the theory of convex optimization in the quaternion domain. For rigorous, this is achieved based on the GHR calculus [24], a generalization of Wirtinger-calculus [25–27] from the complex domain to the quaternion field. Before the introduction of the GHR calculus, the quaternion pseudo-derivative was used for calculating the gradient, which transforms the quaternion optimization problem into a lengthy and complicated real optimization problem; the solution is then found by using real-valued optimization algorithms [28, 29]. Flamant et al. [5, 22] demostrated that the GHR calculus is a powerful theory in quaternion signal processing and non-negative matrix factorization. Mengüç et al. [12, 30] established that the GHR calculus paves the way for the theory and applications of quaternion-valued adaptive filters. Took and Xia [31] proposed a multichannel quaternion least-mean-square based on the GHR calculus for the adaptive filtering. Parcollet et al. [32] further emphasized the significance of the GHR calculus as a recent breakthrough in the field.

The theory of convex optimization in the quaternion field has gained attention due to its promising applications in signal processing and optimization. The aim of this work is to develop the convexity theory of quaternion function using the GHR calculus [24]. To this end, we make use of the duality of the augmented quaternion vector $\boldsymbol{q}_{\mathcal{H}} \triangleq \left(\boldsymbol{q}^{\top}, \boldsymbol{q}^{i\top}, \boldsymbol{q}^{j\top}, \boldsymbol{q}^{k\top}\right)^{\top}$ and the augmented real vector $\boldsymbol{q}_{\mathcal{R}} \triangleq \left(\boldsymbol{q}_a^{\top}, \boldsymbol{q}_b^{\top}, \boldsymbol{q}_c^{\top}, \boldsymbol{q}_d^{\top}\right)^{\top}$ [33]. Next, we employ the relationships between augmented quaternion gradient and augmented real gradient, as well as between the augmented quaternion Hessian matrix and augmented real Hessian matrix, as shown in [33]. Based on these results, we extend the discriminant criteria for convexity from the real field to the augmented quaternion space, $\mathcal{H}$, and then to the quaternion field $\mathbb{H}^n$, as illustrated in Figure 1. Moreover, we define and present four discriminant theorems for strong convexity, by employing the discriminant criteria of convex quaternion functions. Finally, we present a fundamental theorem for the optimality of convex quaternion problems and provide three illustrative applications in the field of signal processing, including quaternion linear mean-square error filter, quaternion projection on affine equality constraint, and quaternion minimum variance beamforming.

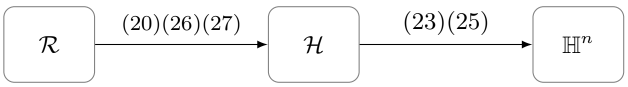

Fig. 1. The derivation process of quaternion optimization theory, with the sets $\mathcal{R}$ and $\mathcal{H}$ defined by (18), (19).

This work makes three significant contributions to the theory of convex optimization in the quaternion field:

- By using the GHR calculus, we establish five discriminant theorems for convex functions in the quaternion field. These theorems include gradient monotonicity and second-order characterization.
- We provide a clear definition and four discriminant criteria for strongly convex functions in the quaternion field; these are consistent with their counterpart real and complex convexity theorems.
- A fundamental theorem is proposed for the optimality of convex quaternion problems, together with some practical applications of convex quaternion optimization in communications and signal processing.

This paper is organized as follows. In Section II, we give an overview of quaternion algebra, the GHR calculus, and some equivalence relationships. Section III presents five discriminant theorems for convex quaternion functions, covering first-order characterization, second-order characterization and some examples of convex quaternion functions. Section IV introduces the definition and discriminant theorems for strongly convex quaternion functions. In Section V, we propose a fundamental theorem for convex quaternion optimization problems and provide three practical applications in signal processing. Finally, this paper concludes with Section VI.

II. Preliminaries

A. Quaternion Algebra

A quaternion, q, can be expressed as

$$q = q_a + q_b i + q_c j + q_d k, \tag{1}$$

where $q_a, q_b, q_c, q_d \in \mathbb{R}$, and the imaginary units i, j and k satisfy $i^2 = j^2 = k^2 = ijk = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. The set of quaternions is defined as $\mathbb{H} \triangleq \{q = q_a + q_b i + q_c j + q_d k \mid q_a, q_b, q_c, q_d \in \mathbb{R}\}$. Owing to the properties of the imaginary units, the multiplication of two quaternions in $\mathbb{H}$ is noncommutative. The real part of q is denoted by $\mathrm{Re}\{q\} = q_a$, whereas the imaginary part (pure quaternion) is $\mathrm{Im}\{q\} = q_b i + q_c j + q_d k$. The conjugate of q is $q^* = \mathrm{Re}\{q\} - \mathrm{Im}\{q\} = q_a - q_b i - q_c j - q_d k$. The modulus of a quaternion is defined as $\mid q \mid = \sqrt{qq^*}$. Define also q^μ, as it is used in Definition 2.1.

Definition 2.1 (Quaternion rotation [34]): For any quaternion, q, and a nonzero quaternion μ, the transformation

$$q^\mu \triangleq \mu q \mu^{-1} \tag{2}$$

describes a rotation of q.

In particular, if μ in (2) is a pure unit quaternion, then the quaternion rotation in (2) becomes quaternion involution [35], such as

$$q^i = -iqi = q_a + iq_b - jq_c - kq_d, \tag{3}$$

$$q^j = -jqj = q_a - iq_b + jq_c - kq_d, \tag{4}$$

$$q^k = -kqk = q_a - iq_b - jq_c + kq_d. \tag{5}$$

Property 2.1 (Properties of quaternion rotation [36]): For any $p, q \in \mathbb{H}$, and $\forall \nu, \mu \in \mathbb{H}$, the following holds

$$\begin{aligned} &(pq)^\mu = p^\mu q^\mu,\ q^{\mu\nu} = (q^\nu)^\mu, \\ &q^{\mu *} \triangleq (q^*)^\mu = (q^\mu)^* \triangleq q^{*\mu}. \end{aligned} \tag{6}$$

B. The GHR Calculus

Definition 2.2 (Real-differentiability [37]): A quaternion function $f : \mathbb{H} \to \mathbb{H}$, given by $f(q) = f_a(q_a, q_b, q_c, q_d) + i f_b(q_a, q_b, q_c, q_d) + j f_c(q_a, q_b, q_c, q_d) + k f_d(q_a, q_b, q_c, q_d)$ is called real differentiable, if f_a, f_b, f_c, f_d are differentiable as functions of the real variables q_a, q_b, q_c, q_d.

Definition 2.3 (GHR derivatives [24]): If $f : \mathbb{H} \to \mathbb{H}$ is real-differentiable, then the left GHR derivatives of the function f with respect to q^μ and $q^{\mu *}$ $(\mu \neq 0, \mu \in \mathbb{H})$ are defined as

$$\frac{\partial f}{\partial q^\mu} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} - \frac{\partial f}{\partial q_b} i^\mu - \frac{\partial f}{\partial q_c} j^\mu - \frac{\partial f}{\partial q_d} k^\mu\right) \in \mathbb{H}, \tag{7}$$

$$\frac{\partial f}{\partial q^{\mu *}} = \frac{1}{4}\left(\frac{\partial f}{\partial q_a} + \frac{\partial f}{\partial q_b} i^\mu + \frac{\partial f}{\partial q_c} j^\mu + \frac{\partial f}{\partial q_d} k^\mu\right) \in \mathbb{H}, \tag{8}$$

where $q = q_a + q_b i + q_c j + q_d k$, $q_a, q_b, q_c, q_d \in \mathbb{R}$, and $\frac{\partial f}{\partial q_a}, \frac{\partial f}{\partial q_b}, \frac{\partial f}{\partial q_c}, \frac{\partial f}{\partial q_d} \in \mathbb{R}$ are the partial derivatives of f with respect to q_a, q_b, q_c, q_d.

Property 2.2 (Properties of the GHR derivatives [24]): If $f : \mathbb{H} \to \mathbb{H}$, $g : \mathbb{H} \to \mathbb{H}$, then

Product rule:

$$\frac{\partial (fg)}{\partial q^\mu} = f\frac{\partial g}{\partial q^\mu} + \frac{\partial f}{\partial q^{g\mu}} g, \quad \frac{\partial (fg)}{\partial q^{\mu *}} = f\frac{\partial g}{\partial q^{\mu *}} + \frac{\partial f}{\partial q^{g\mu *}} g \tag{9}$$

Chain rule:

$$\frac{\partial f(g(q))}{\partial q^\mu} = \sum_{\nu \in \{1,i,j,k\}} \frac{\partial f}{\partial g^\nu}\frac{\partial g^\nu}{\partial q^\mu}, \tag{10}$$

$$\frac{\partial f(g(q))}{\partial q^{\mu *}} = \sum_{\nu \in \{1,i,j,k\}} \frac{\partial f}{\partial g^{\nu *}}\frac{\partial g^{\nu *}}{\partial q^{\mu *}} \tag{11}$$

Rotation rule:

$$\left(\frac{\partial f}{\partial q^\mu}\right)^\nu = \frac{\partial f^\nu}{\partial q^{\nu\mu}}, \quad \left(\frac{\partial f}{\partial q^{\mu *}}\right)^\nu = \frac{\partial f^\nu}{\partial q^{\nu\mu *}} \tag{12}$$

Conjugate rule: If $f : \mathbb{H} \to \mathbb{R}$,

$$\left(\frac{\partial f}{\partial q^\mu}\right)^* = \frac{\partial f}{\partial q^{\mu *}}, \quad \left(\frac{\partial f}{\partial q^{\mu *}}\right)^* = \frac{\partial f}{\partial q^\mu}. \tag{13}$$

Definition 2.4 (Quaternion gradient [33]): The quaternion gradient and its conjugate gradient of a function $f : \mathbb{H}^n \to \mathbb{R}$ are defined as

$$\nabla_{\boldsymbol{q}} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}}\right)^{\mathsf{T}} = \left(\frac{\partial f}{\partial q_1}, \ldots, \frac{\partial f}{\partial q_n}\right)^{\mathsf{T}} \in \mathbb{H}^n, \quad (14)$$

$$\nabla_{\boldsymbol{q}^*} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}^*}\right)^{\mathsf{T}} = \left(\frac{\partial f}{\partial q_1^*}, \ldots, \frac{\partial f}{\partial q_n^*}\right)^{\mathsf{T}} \in \mathbb{H}^n, \quad (15)$$

where $\left(\frac{\partial f}{\partial \boldsymbol{q}}\right)^{\mathsf{T}}$ is the transpose of $\frac{\partial f}{\partial \boldsymbol{q}}$.

Definition 2.5 (Quaternion Hessian [33]): Let $f : \mathbb{H}^n \to \mathbb{R}$, then the two quaternion Hessian matrices are defined as

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}} &\triangleq \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{\partial f}{\partial \boldsymbol{q}}\right)^{\mathsf{T}} \\ &= \begin{pmatrix} \frac{\partial^2 f}{\partial q_1 \partial q_1} & \cdots & \frac{\partial^2 f}{\partial q_n \partial q_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial q_1 \partial q_n} & \cdots & \frac{\partial^2 f}{\partial q_n \partial q_n} \end{pmatrix} \in \mathbb{H}^{n\times n}, \quad (16)\end{aligned}$$

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} &\triangleq \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{\partial f}{\partial \boldsymbol{q}^*}\right)^{\mathsf{T}} \\ &= \begin{pmatrix} \frac{\partial^2 f}{\partial q_1 \partial q_1^*} & \cdots & \frac{\partial^2 f}{\partial q_n \partial q_1^*} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial q_1 \partial q_n^*} & \cdots & \frac{\partial^2 f}{\partial q_n \partial q_n^*} \end{pmatrix} \in \mathbb{H}^{n\times n}. \quad (17)\end{aligned}$$

C. The Relationship of Augmented Quaternion and the Augmented Real Vector, Gradient, and Hessian Matrix

Consider a quaternion vector $\boldsymbol{q} = \boldsymbol{q}_a + \boldsymbol{q}_b i + \boldsymbol{q}_c j + \boldsymbol{q}_d k \in \mathbb{H}^n$ where $\boldsymbol{q}_a, \boldsymbol{q}_b, \boldsymbol{q}_c, \boldsymbol{q}_d \in \mathbb{R}^n$. Define its augmented real vector as $\boldsymbol{q}_{\mathcal{R}} \triangleq \left(\boldsymbol{q}_a^{\mathsf{T}}, \boldsymbol{q}_b^{\mathsf{T}}, \boldsymbol{q}_c^{\mathsf{T}}, \boldsymbol{q}_d^{\mathsf{T}}\right)^{\mathsf{T}} \in \mathcal{R}$ [4, 38] and the augmented quaternion vector as $\boldsymbol{q}_{\mathcal{H}} \triangleq \left(\boldsymbol{q}^{\mathsf{T}}, \boldsymbol{q}^{i\mathsf{T}}, \boldsymbol{q}^{j\mathsf{T}}, \boldsymbol{q}^{k\mathsf{T}}\right)^{\mathsf{T}} \in \mathcal{H}$ [33], where the set of augmented real vectors and the set of augmented quaternion vectors are defined as

$$\mathcal{R} \triangleq \left\{\boldsymbol{q}_{\mathcal{R}} = \left(\boldsymbol{q}_a^{\mathsf{T}}, \boldsymbol{q}_b^{\mathsf{T}}, \boldsymbol{q}_c^{\mathsf{T}}, \boldsymbol{q}_d^{\mathsf{T}}\right)^{\mathsf{T}} \mid \boldsymbol{q} \in \mathbb{H}^n\right\} = \mathbb{R}^{4n}, \quad (18)$$

$$\mathcal{H} \triangleq \left\{\boldsymbol{q}_{\mathcal{H}} = \left(\boldsymbol{q}^{\mathsf{T}}, \boldsymbol{q}^{i\mathsf{T}}, \boldsymbol{q}^{j\mathsf{T}}, \boldsymbol{q}^{k\mathsf{T}}\right)^{\mathsf{T}} \mid \boldsymbol{q} \in \mathbb{H}^n\right\} \subset \mathbb{H}^{4n}. \quad (19)$$

By definition, there exists a one-to-one mapping between $\mathbb{H}^n$, $\mathcal{R}$ and $\mathcal{H}$ [22].

Proposition 2.1 ([33]): The relationship between the augmented quaternion vector, $\boldsymbol{q}_{\mathcal{H}}$, and the augmented real vector, $\boldsymbol{q}_{\mathcal{R}}$, is given by

$$\boldsymbol{q}_{\mathcal{H}} = \boldsymbol{J}_n \boldsymbol{q}_{\mathcal{R}} \quad \Leftrightarrow \quad \boldsymbol{q}_{\mathcal{R}} = \frac{1}{4}\boldsymbol{J}_n^{\mathsf{H}} \boldsymbol{q}_{\mathcal{H}}, \quad (20)$$

where

$$\boldsymbol{J}_n = \begin{pmatrix} \boldsymbol{I}_n & i\boldsymbol{I}_n & j\boldsymbol{I}_n & k\boldsymbol{I}_n \\ \boldsymbol{I}_n & i\boldsymbol{I}_n & -j\boldsymbol{I}_n & -k\boldsymbol{I}_n \\ \boldsymbol{I}_n & -i\boldsymbol{I}_n & j\boldsymbol{I}_n & -k\boldsymbol{I}_n \\ \boldsymbol{I}_n & -i\boldsymbol{I}_n & -j\boldsymbol{I}_n & k\boldsymbol{I}_n \end{pmatrix} \in \mathbb{H}^{4n\times 4n}, \quad (21)$$

and $\boldsymbol{J}_n^{\mathsf{H}} \boldsymbol{J}_n = 4\boldsymbol{I}_{4n}$, while $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, with $\boldsymbol{J}_n^{\mathsf{H}}$ as the conjugate transpose of $\boldsymbol{J}_n$.

From (20), the quaternion function $f(\boldsymbol{q}) : \mathbb{H}^n \to \mathbb{R}$ can be viewed in three equivalent forms [33], as follows

$$\begin{aligned} f(\boldsymbol{q}) \quad &\Leftrightarrow \quad f(\boldsymbol{q}_{\mathcal{R}}) \triangleq f(\boldsymbol{q}_a, \boldsymbol{q}_b, \boldsymbol{q}_c, \boldsymbol{q}_d) \\ &\Leftrightarrow \quad f(\boldsymbol{q}_{\mathcal{H}}) \triangleq f\left(\boldsymbol{q}, \boldsymbol{q}^i, \boldsymbol{q}^j, \boldsymbol{q}^k\right). \end{aligned} \quad (22)$$

Note that these three functions are equivalent but have different forms, denoted as f for simplicity. Here, the variables of the functions $f(\boldsymbol{q})$, $f(\boldsymbol{q}_{\mathcal{R}})$, and $f(\boldsymbol{q}_{\mathcal{H}})$ are quaternion vectors, augmented real vectors, and augmented quaternion vectors, respectively. They are referred to as quaternion function, augmented real function, and augmented quaternion function, respectively. For $f(\boldsymbol{q}_{\mathcal{R}}) : \mathcal{R} \to \mathbb{R}$, its augmented real gradient is defined as $\nabla_{\mathcal{R}} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{R}}}\right)^{\mathsf{T}}$ and the augmented real Hessian matrix as $\boldsymbol{H}_{\mathcal{R}\mathcal{R}} \triangleq \frac{\partial}{\partial \boldsymbol{q}_{\mathcal{R}}}\left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{R}}}\right)^{\mathsf{T}}$. For $f(\boldsymbol{q}_{\mathcal{H}}) : \mathcal{H} \to \mathbb{R}$, the augmented quaternion gradient and its conjugate gradient are defined as [39]

$$\nabla_{\mathcal{H}} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{H}}}\right)^{\mathsf{T}} = \begin{pmatrix} \nabla_{\boldsymbol{q}} f \\ \nabla_{\boldsymbol{q}^i} f \\ \nabla_{\boldsymbol{q}^j} f \\ \nabla_{\boldsymbol{q}^k} f \end{pmatrix}, \quad (23)$$

$$\nabla_{\mathcal{H}^*} f \triangleq \left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{H}}^*}\right)^{\mathsf{T}} = \begin{pmatrix} \nabla_{\boldsymbol{q}^*} f \\ \nabla_{\boldsymbol{q}^{i*}} f \\ \nabla_{\boldsymbol{q}^{j*}} f \\ \nabla_{\boldsymbol{q}^{k*}} f \end{pmatrix}, \quad (24)$$

and the augmented quaternion Hessian matrix is defined as

$$\begin{aligned}\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} &\triangleq \frac{\partial}{\partial \boldsymbol{q}_{\mathcal{H}}}\left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{H}}^*}\right)^{\mathsf{T}} \quad (25) \\ &= \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^*} \\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{i*}} \\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{j*}} \\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{k*}} \end{pmatrix}.\end{aligned}$$

Proposition 2.2 ([39]): The relationship between the augmented quaternion gradient, $\nabla_{\mathcal{H}^*} f$, and the augmented real gradient, $\nabla_{\mathcal{R}} f$, is given by

$$\nabla_{\mathcal{H}^*} f = \frac{1}{4}\boldsymbol{J}_n \nabla_{\mathcal{R}} f \quad \Leftrightarrow \quad \nabla_{\mathcal{R}} f = \boldsymbol{J}_n{}^{\mathsf{H}} \nabla_{\mathcal{H}^*} f. \quad (26)$$

Proposition 2.3 ([33]): The relationship between the augmented quaternion Hessian matrix, $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$, and the augmented real Hessian matrix, $\boldsymbol{H}_{\mathcal{R}\mathcal{R}}$, is given by

$$\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} = \frac{1}{16}\boldsymbol{J}_n \boldsymbol{H}_{\mathcal{R}\mathcal{R}} \boldsymbol{J}_n^{\mathsf{H}} \;\Leftrightarrow\; \boldsymbol{H}_{\mathcal{R}\mathcal{R}} = \boldsymbol{J}_n^{\mathsf{H}} \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \boldsymbol{J}_n \quad (27)$$

where $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \triangleq \frac{\partial}{\partial \boldsymbol{q}_{\mathcal{H}}}\left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{H}}^*}\right)^{\mathsf{T}}$, $\boldsymbol{H}_{\mathcal{R}\mathcal{R}} \triangleq \frac{\partial}{\partial \boldsymbol{q}_{\mathcal{R}}}\left(\frac{\partial f}{\partial \boldsymbol{q}_{\mathcal{R}}}\right)^{\mathsf{T}}$.

Corollary 2.1: The augmented quaternion Hessian matrix, $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$, is a Hermite matrix, that is

$$\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}^{\mathsf{H}} = \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}, \quad (28)$$

where $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}^{\mathsf{H}}$ is the conjugate transpose of $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$.

Proof: This is straightforward to demonstrate by using (27) and the fact that $\boldsymbol{H}_{\mathcal{R}\mathcal{R}}$ is a Hermitian matrix. ■

Proposition 2.4: For any $\boldsymbol{p},\boldsymbol{q}\in\mathbb{H}^n$, their augmented real vectors are $\boldsymbol{p}_{\mathcal{R}},\boldsymbol{q}_{\mathcal{R}}\in\mathcal{R}$, and their augmented quaternion vectors are $\boldsymbol{p}_{\mathcal{H}},\boldsymbol{q}_{\mathcal{H}}\in\mathcal{H}$. Then

$$(a)\quad \boldsymbol{p}_{\mathcal{H}}^{\mathsf{T}}\boldsymbol{q}_{\mathcal{H}}=4\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{T}}\boldsymbol{q}\right\}; \tag{29}$$
$$(b)\quad 4\boldsymbol{p}_{\mathcal{R}}^{\mathsf{T}}\boldsymbol{q}_{\mathcal{R}}=\boldsymbol{p}_{\mathcal{H}}^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}=4\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}\right\}; \tag{30}$$
$$(c)\quad 2\|\boldsymbol{p}_{\mathcal{R}}\|_2=\|\boldsymbol{p}_{\mathcal{H}}\|_2=2\|\boldsymbol{p}\|_2; \tag{31}$$
$$(d)\quad \|\boldsymbol{p}+\boldsymbol{q}\|_2^2=\|\boldsymbol{p}\|_2^2+2\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}\right\}+\|\boldsymbol{q}\|_2^2. \tag{32}$$

Proof: By the relationship of $\boldsymbol{q}$, $\boldsymbol{q}_{\mathcal{R}}$, and $\boldsymbol{q}_{\mathcal{H}}$, we have

$$\begin{aligned}(a)\ \boldsymbol{p}_{\mathcal{H}}^{\mathsf{T}}\boldsymbol{q}_{\mathcal{H}}&=\sum_{\mu\in\{1,i,j,k\}}\boldsymbol{p}^{\mu\mathsf{T}}\boldsymbol{q}^{\mu}\overset{(6)}{=}\sum_{\mu\in\{1,i,j,k\}}\left(\boldsymbol{p}^{\mathsf{T}}\boldsymbol{q}\right)^{\mu}\\&\overset{(20)}{=}4\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{T}}\boldsymbol{q}\right\};\end{aligned} \tag{33}$$

$$\begin{aligned}(b)\ 4\boldsymbol{p}_{\mathcal{R}}^{\mathsf{T}}\boldsymbol{q}_{\mathcal{R}}&=4\boldsymbol{p}_{\mathcal{R}}^{\mathsf{H}}\boldsymbol{q}_{\mathcal{R}}\overset{(20)}{=}\boldsymbol{p}_{\mathcal{H}}^{\mathsf{H}}\boldsymbol{J}_n\frac{1}{4}\boldsymbol{J}_n^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}=\boldsymbol{p}_{\mathcal{H}}^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}\\&\overset{(29)}{=}4\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}\right\};\end{aligned} \tag{34}$$

$$\begin{aligned}(c)\ 4\|\boldsymbol{p}_{\mathcal{R}}\|_2^2&=4\boldsymbol{p}_{\mathcal{R}}^{\mathsf{T}}\boldsymbol{p}_{\mathcal{R}}\overset{(30)}{=}\boldsymbol{p}_{\mathcal{H}}^{\mathsf{H}}\boldsymbol{p}_{\mathcal{H}}=\|\boldsymbol{p}_{\mathcal{H}}\|_2^2\\&\overset{(30)}{=}4\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{p}\right\}=4\boldsymbol{p}^{\mathsf{H}}\boldsymbol{p}=4\|\boldsymbol{p}\|_2^2;\end{aligned} \tag{35}$$

$$\begin{aligned}(d)\ \|\boldsymbol{p}+\boldsymbol{q}\|_2^2&=(\boldsymbol{p}+\boldsymbol{q})^{\mathsf{H}}(\boldsymbol{p}+\boldsymbol{q})\\&=\boldsymbol{p}^{\mathsf{H}}\boldsymbol{p}+\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}+\boldsymbol{q}^{\mathsf{H}}\boldsymbol{p}+\boldsymbol{q}^{\mathsf{H}}\boldsymbol{q}\\&=\|\boldsymbol{p}\|_2^2+2\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}\right\}+\|\boldsymbol{q}\|_2^2.\end{aligned} \tag{36}$$

This completes the proof. ■

Proposition 2.5: If the quaternion function $f(\boldsymbol{q}):\mathbb{H}^n\to\mathbb{R}$ is real-differentiable, then $\forall\boldsymbol{p},\boldsymbol{q}\in\mathbb{H}^n$ we have

$$\begin{aligned}(a)\ \nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{T}}\boldsymbol{q}_{\mathcal{R}}&=\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}\\&=4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\boldsymbol{q}\right\};\end{aligned} \tag{37}$$

$$\begin{aligned}(b)\ \nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{T}}\nabla_{\mathcal{R}}f(\boldsymbol{q}_{\mathcal{R}})&=4\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\nabla_{\mathcal{H}^*}f(\boldsymbol{q}_{\mathcal{H}})\\&=16\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q})\right\};\end{aligned} \tag{38}$$

$$(c)\ \|\nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})\|_2=2\|\nabla_{\mathcal{H}}f(\boldsymbol{p}_{\mathcal{H}})\|_2=4\|\nabla_{\boldsymbol{p}}f(\boldsymbol{p})\|_2. \tag{39}$$

Proof: By the relationship of $\boldsymbol{q}$, $\boldsymbol{q}_{\mathcal{R}}$, and $\boldsymbol{q}_{\mathcal{H}}$, and the relationship of $\nabla_{\boldsymbol{q}^*}f$, $\nabla_{\mathcal{R}}f$, and $\nabla_{\mathcal{H}^*}f$, we have

$$\begin{aligned}(a)\ \nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{T}}\boldsymbol{q}_{\mathcal{R}}&=\nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{H}}\boldsymbol{q}_{\mathcal{R}}\\&\overset{(20)(26)}{=}\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\boldsymbol{J}_n\frac{1}{4}\boldsymbol{J}_n^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}\\&=\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\boldsymbol{q}_{\mathcal{H}}\\&=\sum_{\mu\in\{1,i,j,k\}}\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mu\mathsf{H}}\boldsymbol{q}^{\mu}\\&\overset{(6)}{=}\sum_{\mu\in\{1,i,j,k\}}\left(\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\boldsymbol{q}\right)^{\mu}\\&\overset{(20)}{=}4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\boldsymbol{q}\right\};\end{aligned} \tag{40}$$

$$\begin{aligned}(b)\ \nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{T}}\nabla_{\mathcal{R}}f(\boldsymbol{q}_{\mathcal{R}})&=\nabla_{\mathcal{R}}f(\boldsymbol{p}_{\mathcal{R}})^{\mathsf{H}}\nabla_{\mathcal{R}}f(\boldsymbol{q}_{\mathcal{R}})\\&\overset{(26)}{=}\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\boldsymbol{J}_n\boldsymbol{J}_n^{\mathsf{H}}\nabla_{\mathcal{H}^*}f(\boldsymbol{q}_{\mathcal{H}})\\&=4\nabla_{\mathcal{H}^*}f(\boldsymbol{p}_{\mathcal{H}})^{\mathsf{H}}\nabla_{\mathcal{H}^*}f(\boldsymbol{q}_{\mathcal{H}})\\&=4\sum_{\mu\in\{1,i,j,k\}}\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mu\mathsf{H}}\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q})^{\mu}\\&\overset{(6)}{=}4\sum_{\mu\in\{1,i,j,k\}}\left(\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q})\right)^{\mu}\\&\overset{(20)}{=}16\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q})\right\};\end{aligned} \tag{41}$$

(c) Let $\boldsymbol{q}=\boldsymbol{p}$ in (38).

This completes the proof. ■

III. Discriminant Theorems for Convex Quaternion Functions

The objective of this section is to introduce five discriminant criteria for convex quaternion functions, including the first-order characterization and the second-order characterization. An example is presented to illustrate how these criteria can be applied in practice.

A. Convex Set and Convex Quaternion Function

We begin by introducing the fundamental concepts, such as convex set and convex function [40, 41].

Definition 3.1 (Convex set): The set $\mathcal{C}$ is called convex, if $\forall\boldsymbol{x},\boldsymbol{y}\in\mathcal{C}$, $\forall 0\leqslant\theta\leqslant 1$, $\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\in\mathcal{C}$. The set $\mathcal{C}$ can be a subset of $\mathbb{H}^n$, $\mathcal{R}$ or $\mathcal{H}$.

Definition 3.2 (Convex function): A function f is said to be convex, if $\boldsymbol{dom}f$ is convex, and $\forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{dom}f$, $0\leqslant\theta\leqslant 1$,

$$f\big(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\big)\leqslant\theta f(\boldsymbol{x})+(1-\theta)f(\boldsymbol{y}). \tag{42}$$

The range of the function f is $\mathbb{R}$, and the definition field $\boldsymbol{dom}f$ can be a subset of $\mathbb{H}^n$, $\mathcal{R}$ or $\mathcal{H}$.

Example 3.1: Consider a quaternion matrix, $\boldsymbol{A}\in\mathbb{H}^{m\times n}$, and a quaternion vector, $\boldsymbol{b}\in\mathbb{H}^m$, then the set $\mathcal{D}\triangleq\{\boldsymbol{q}\in\mathbb{H}^n\mid\boldsymbol{A}\boldsymbol{q}=\boldsymbol{b}\}$ is convex.

Proof: $\forall\boldsymbol{p},\boldsymbol{q}\in\mathcal{D}$, $\boldsymbol{A}\boldsymbol{p}=\boldsymbol{b}$, $\boldsymbol{A}\boldsymbol{q}=\boldsymbol{b}$, $\forall 0\leqslant\theta\leqslant 1$,

$$\boldsymbol{A}(\theta\boldsymbol{p}+(1-\theta)\boldsymbol{q})=\theta\boldsymbol{A}\boldsymbol{p}+(1-\theta)\boldsymbol{A}\boldsymbol{q}=\theta\boldsymbol{b}+(1-\theta)\boldsymbol{b}=\boldsymbol{b}. \tag{43}$$

Therefore, $\theta\boldsymbol{p}+(1-\theta)\boldsymbol{q}\in\mathcal{D}$, that is the set $\mathcal{D}$ is convex. ■

Example 3.2: If the quaternion function $f(\boldsymbol{q})$ is convex, then the set $\mathcal{E}\triangleq\{\boldsymbol{q}\in\mathbb{H}^n\mid f(\boldsymbol{q})\leqslant 0\}$ is also convex.

Proof: $\forall\boldsymbol{p},\boldsymbol{q}\in\mathcal{E}$, $f(\boldsymbol{p})\leqslant 0$, $f(\boldsymbol{q})\leqslant 0$. Since $f(\boldsymbol{q})$ is convex, $\forall 0\leqslant\theta\leqslant 1$,

$$f\big(\theta\boldsymbol{p}+(1-\theta)\boldsymbol{q}\big)\leqslant\theta f(\boldsymbol{p})+(1-\theta)f(\boldsymbol{q})\leqslant 0. \tag{44}$$

Therefore, $\theta\boldsymbol{p}+(1-\theta)\boldsymbol{q}\in\mathcal{E}$, that is the set $\mathcal{E}$ is convex. ■

B. First-order Characterization of Discriminant Theorems for Convex Quaternion Functions

We shall now introduce four discriminant theorems for convex quaternion functions, including the first-order characterization and gradient monotonicity.

Theorem 3.1: Consider the three sets $\mathcal{C}\subset\mathbb{H}^n$,
$\mathcal{C}_{\mathcal{R}}\triangleq\left\{\boldsymbol{q}_{\mathcal{R}}=\left(\boldsymbol{q}_a^{\mathsf{T}},\boldsymbol{q}_b^{\mathsf{T}},\boldsymbol{q}_c^{\mathsf{T}},\boldsymbol{q}_d^{\mathsf{T}}\right)^{\mathsf{T}}\mid\boldsymbol{q}\in\mathcal{C}\right\}\subset\mathcal{R}=\mathbb{R}^{4n}$,

$\mathcal{C}_\mathcal{H} \triangleq \left\{ \boldsymbol{q}_\mathcal{H} = \left(\boldsymbol{q}^\mathsf{T}, \boldsymbol{q}^{i\mathsf{T}}, \boldsymbol{q}^{j\mathsf{T}}, \boldsymbol{q}^{k\mathsf{T}} \right)^\mathsf{T} \mid \boldsymbol{q} \in \mathcal{C} \right\} \subset \mathcal{H} \subset \mathbb{H}^{4n}$. Then, $\mathcal{C}$ is convex $\Leftrightarrow \mathcal{C}_\mathcal{R}$ is convex $\Leftrightarrow \mathcal{C}_\mathcal{H}$ is convex.

Proof: Using the definition of $\mathcal{C}$, $\mathcal{C}_\mathcal{R}$, $\mathcal{C}_\mathcal{H}$, and that of convex set, the proof following. ■

A straightforward method to discriminate the convexity of a quaternion function is to confine it to a line segment and determine whether the resulting one-dimensional function is convex, as in the following theorem.

Theorem 3.2: The quaternion function $f(\boldsymbol{q}) : \mathcal{C} \subset \mathbb{H}^n \to \mathbb{R}$ is convex if and only if (shortened to iff) $\forall \boldsymbol{q} \in \mathcal{C}$, $\boldsymbol{v} \in \mathbb{H}^n$, $g : \mathcal{S} \to \mathbb{R}$,

$$g(t) = f(\boldsymbol{q} + t\boldsymbol{v}) \tag{45}$$

is convex, where $\mathcal{S} \triangleq \{t \in \mathbb{R} \mid \boldsymbol{q} + t\boldsymbol{v} \in \mathcal{C}\} \subset \mathbb{R}$.

Proof: The proof follows the same steps as its counterpart in the real field [40, 41]. ■

For real-differentiable quaternion functions, we can also use their gradient information to discriminate their convexity, as stated in the following theorem.

Theorem 3.3 (First-order characterization [22]): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then $f(\boldsymbol{q})$ is convex iff $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$f(\boldsymbol{q}) \geqslant f(\boldsymbol{p}) + 4\mathrm{Re}\left\{ \nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})^\mathsf{H} (\boldsymbol{q} - \boldsymbol{p}) \right\}, \tag{46}$$

where $\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})$ is defined in (15).

Another commonly used first-order characterization is gradient monotonicity, as shown below.

Theorem 3.4 (Gradient monotonicity): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, $f(\boldsymbol{q})$ is convex iff $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$\mathrm{Re}\left\{ \left(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q}) \right)^\mathsf{H} (\boldsymbol{p} - \boldsymbol{q}) \right\} \geqslant 0, \tag{47}$$

where $\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})$ is defined in (15).

Proof: From Theorem 3.1, $\mathcal{C}$ is convex iff $\mathcal{C}_\mathcal{R}$ is convex. We already know [40, 41] that for a differentiable real function, $f(\boldsymbol{q}_\mathcal{R})$ is convex iff $\forall \boldsymbol{p}_\mathcal{R}, \boldsymbol{q}_\mathcal{R} \in \mathcal{C}_\mathcal{R}$,

$$\left(\nabla_\mathcal{R} f(\boldsymbol{p}_\mathcal{R}) - \nabla_\mathcal{R} f(\boldsymbol{q}_\mathcal{R}) \right)^\mathsf{T} (\boldsymbol{p}_\mathcal{R} - \boldsymbol{q}_\mathcal{R}) \geqslant 0, \tag{48}$$

where the set $\mathcal{C}_\mathcal{R} \subset \mathcal{R}$ is convex. Hence from (37), we have

$$\begin{aligned} &\left(\nabla_\mathcal{R} f(\boldsymbol{p}_\mathcal{R}) - \nabla_\mathcal{R} f(\boldsymbol{q}_\mathcal{R}) \right)^\mathsf{T} (\boldsymbol{p}_\mathcal{R} - \boldsymbol{q}_\mathcal{R}) \\ =&4\mathrm{Re}\left\{ \left(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q}) \right)^\mathsf{H} (\boldsymbol{p} - \boldsymbol{q}) \right\}. \end{aligned} \tag{49}$$

Upon substituting (49) into (48), the proof follows. ■

In addition, we can also use the epigraph to discriminate the convexity of $f(\boldsymbol{q})$, as shown below.

Definition 3.3 (Epigraph): For the quaternion generalized real-valued function $f(\boldsymbol{q}) : \mathbb{H}^n \to \mathbb{R} \cup \{\pm\infty\}$, the set

$$\boldsymbol{epi} f = \left\{ (\boldsymbol{q}, t) \in \mathbb{H}^{n+1} \mid f(\boldsymbol{q}) \leqslant t, t \in \mathbb{R} \right\} \tag{50}$$

is called the epigraph of $f(\boldsymbol{q})$.

Theorem 3.5: The quaternion generalized real-valued function $f(\boldsymbol{q}) : \mathcal{C} \subset \mathbb{H}^n \to \mathbb{R} \cup \{\pm\infty\}$ is convex, iff $\boldsymbol{epi} f$ is a convex set.

Proof: The proof follows the same steps as its counterpart in the real field [40, 41]. ■

C. Second-order Characterization of Discriminant Theorems for Convex Quaternion Functions

Before introducing the second-order characterization of convex quaternion functions, we first need to define positive definite quaternion matrices.

Definition 3.4 (Positive definite matrix): The matrix $\boldsymbol{A} \in \mathbb{H}^{n\times n}$ is called positive definite, if

$$\mathrm{Re}\left\{ \boldsymbol{x}^\mathsf{H} \boldsymbol{A} \boldsymbol{x} \right\} > 0, \quad \forall \boldsymbol{x} \in \mathbb{H}^n, \boldsymbol{x} \neq \boldsymbol{0}, \tag{51}$$

and is denoted by $\boldsymbol{A} \succ \boldsymbol{O}$, where $\boldsymbol{O}$ is the $n \times n$ zero matrix. Similarly, the matrix $\boldsymbol{A} \in \mathbb{H}^{n\times n}$ is called positive semi-definite, if

$$\mathrm{Re}\left\{ \boldsymbol{x}^\mathsf{H} \boldsymbol{A} \boldsymbol{x} \right\} \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{H}^n, \boldsymbol{x} \neq \boldsymbol{0}, \tag{52}$$

and is denoted by $\boldsymbol{A} \succeq \boldsymbol{O}$.

Theorem 3.6: If the matrix $\boldsymbol{A} \in \mathbb{H}^{n\times n}$ satisfies $\boldsymbol{A}^\mathsf{H} = \boldsymbol{A}$, then $\boldsymbol{A}$ is positive definite iff

$$\boldsymbol{x}^\mathsf{H} \boldsymbol{A} \boldsymbol{x} > 0, \quad \forall \boldsymbol{x} \in \mathbb{H}^n, \boldsymbol{x} \neq \boldsymbol{0}. \tag{53}$$

Similarly, the matrix $\boldsymbol{A}$ is positive semi-definite iff

$$\boldsymbol{x}^\mathsf{H} \boldsymbol{A} \boldsymbol{x} \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{H}^n, \boldsymbol{x} \neq \boldsymbol{0}. \tag{54}$$

Proof: This is straightforward to prove, by applying Definition 3.4. ■

If the quaternion function $f(\boldsymbol{q})$ is second-order continuous real-differentiable, we can use the Hessian matrix to discriminate its convexity, as shown below.

Theorem 3.7 (Second-order characterization): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then $f(\boldsymbol{q})$ is convex iff

$$\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \boldsymbol{O}. \tag{55}$$

where $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$ is defined in (25).

Proof: Applying Theorem 3.1, the set $\mathcal{C}$ is convex iff the set $\mathcal{C}_\mathcal{R}$ is convex. We already know [40, 41] that for a second-order continous differentiable function, $f(\boldsymbol{q}_\mathcal{R})$ is convex iff

$$\boldsymbol{H}_{\mathcal{R}\mathcal{R}} \succeq \boldsymbol{O}, \quad \forall \boldsymbol{q}_\mathcal{R} \in \mathcal{C}_\mathcal{R}, \tag{56}$$

where the set $\mathcal{C}_\mathcal{R} \subset \mathcal{R}$ is convex. By Corollary 2.1, $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$ is a Hermite matrix. Then $\forall \boldsymbol{x}_\mathcal{H} \in \mathcal{H}$, $\boldsymbol{x}_\mathcal{H} \neq \boldsymbol{0}$, we have

$$\begin{aligned} \boldsymbol{x}_\mathcal{H}^\mathsf{H} \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \boldsymbol{x}_\mathcal{H} &\overset{(27)}{=} \frac{1}{16} \boldsymbol{x}_\mathcal{H}^\mathsf{H} \boldsymbol{J}_n \boldsymbol{H}_{\mathcal{R}\mathcal{R}} \boldsymbol{J}_n^\mathsf{H} \boldsymbol{x}_\mathcal{H} \\ &= \frac{1}{16} \left(\boldsymbol{J}_n^\mathsf{H} \boldsymbol{x}_\mathcal{H} \right)^\mathsf{H} \boldsymbol{H}_{\mathcal{R}\mathcal{R}} \left(\boldsymbol{J}_n^\mathsf{H} \boldsymbol{x}_\mathcal{H} \right) \\ &\overset{(20)}{=} \boldsymbol{x}_\mathcal{R}^\mathsf{H} \boldsymbol{H}_{\mathcal{R}\mathcal{R}} \boldsymbol{x}_\mathcal{R}. \end{aligned} \tag{57}$$

Therefore,

$$\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \boldsymbol{O} \quad \Leftrightarrow \quad \boldsymbol{H}_{\mathcal{R}\mathcal{R}} \succeq \boldsymbol{O}, \tag{58}$$

which concludes the proof. ■

Corollary 3.1: Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, the following three propositions are equivalent:

(a) $f(\boldsymbol{q})$ is convex;

(b) $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \boldsymbol{O}$;

$$(c)\quad \sum_{\nu\in\{1,i,j,k\}} \mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^{*}}\boldsymbol{x}^{\nu}\right\}\geqslant 0,\ \ \forall \boldsymbol{x}\in\mathbb{H}^{n},\, \boldsymbol{x}\neq \mathbf{0}.$$

Proof: From Theorem 3.7, (a) is equivalent to (b), so we only need to prove that (b) is equivalent to (c). From Corollary 2.1, we know that $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$ is a Hermite matrix. Then $\forall \boldsymbol{x}_{\mathcal{H}}\in\mathcal{H}$, $\boldsymbol{x}_{\mathcal{H}}\neq\mathbf{0}$, we have

$$\begin{aligned}
&\boldsymbol{x}_{\mathcal{H}}^{\mathsf{H}}\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}\boldsymbol{x}_{\mathcal{H}}\\
&=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}^{i}\\ \boldsymbol{x}^{j}\\ \boldsymbol{x}^{k}\end{pmatrix}^{\mathsf{H}}\begin{pmatrix}\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^*} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^*}\\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{i*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{i*}}\\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{j*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{j*}}\\ \boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^i\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^j\boldsymbol{q}^{k*}} & \boldsymbol{H}_{\boldsymbol{q}^k\boldsymbol{q}^{k*}}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}^{i}\\ \boldsymbol{x}^{j}\\ \boldsymbol{x}^{k}\end{pmatrix}\\
&=\sum_{\mu,\nu\in\{1,i,j,k\}}\boldsymbol{x}^{\mu\mathsf{H}}\boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^{\mu*}}\boldsymbol{x}^{\nu}\\
&\overset{(20)}{=}4\sum_{\nu\in\{1,i,j,k\}}\mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^{*}}\boldsymbol{x}^{\nu}\right\}.
\end{aligned}\tag{59}$$

Therefore,

$$\begin{aligned}
&\sum_{\nu\in\{1,i,j,k\}}\mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^{*}}\boldsymbol{x}^{\nu}\right\}\geqslant 0,\quad \forall\boldsymbol{x}\in\mathbb{H}^{n},\,\boldsymbol{x}\neq\mathbf{0}\\
&\qquad\Leftrightarrow\quad \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}\succeq\boldsymbol{O}.
\end{aligned}\tag{60}$$

This completes the proof. ■

Lemma 3.1: The matrix $\boldsymbol{A}\in\mathbb{H}^{n\times n}$ is positive definite (positive semi-definite), iff all principal submatrices of $\boldsymbol{A}$ are positive definite (positive semi-definite).

Proof: The follows the same steps as its counterpart in the real field [42]. ■

Applying Lemma 3.1, we can obtain a necessary condition for convex quaternion functions.

Theorem 3.8: Consider a convex set $\mathcal{C}\subset\mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}):\mathcal{C}\to\mathbb{R}$. If $f(\boldsymbol{q})$ is convex, then

$$\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*}\succeq\boldsymbol{O},\tag{61}$$

where $\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*}$ is the quaternion Hessian matrix, defined in (17).

Proof: Upon applying Theorem 3.7, together with the convexity of $f(\boldsymbol{q})$, we have $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}\succeq\boldsymbol{O}$. By (25) and Lemma 3.1, we finally obtain $\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*}\succeq\boldsymbol{O}$. ■

D. Examples of Convex Quaternion Function

We next provide a basic example to determine the convexity of quaternion functions. In this example, we make use of certain GHR derivatives presented in TABLE IV of [36], which are included in TABLE *I* here.

Example 3.3: If the quaternion function $f(\boldsymbol{q})=\|\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b}\|_2^2$, $\forall\boldsymbol{q}\in\mathbb{H}^n$, $\boldsymbol{A}\in\mathbb{H}^{m\times n}$, $\boldsymbol{b}\in\mathbb{H}^m$, then $f(\boldsymbol{q})$ is convex.

Proof: (**First-order characterization criterion**) By the definition of the 2-norm, we have

$$\begin{aligned}
f(\boldsymbol{q})=&\|\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b}\|_2^2\\
=&(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})\\
=&\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}-\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{b}-\boldsymbol{b}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}+\boldsymbol{b}^{\mathsf{H}}\boldsymbol{b}.
\end{aligned}\tag{62}$$

TABLE I
SEVERAL DERIVATIVES PERFORMED BY THE GHR CALCULUS FROM TABLE IV OF [36], $\forall\boldsymbol{A}\in\mathbb{H}^{n\times n}$, $\forall\boldsymbol{a}\in\mathbb{H}^n$, $\forall\boldsymbol{b}\in\mathbb{H}^n$, $\alpha\in\mathbb{H}$, $\beta\in\mathbb{H}$.

$f(\boldsymbol{q})$ or $\boldsymbol{f}(\boldsymbol{q})$	$\frac{\partial f}{\partial \boldsymbol{q}}$ or $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}}$	$\frac{\partial f}{\partial \boldsymbol{q}^*}$ or $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}^*}$
$\boldsymbol{a}^{\mathsf{T}}\boldsymbol{q}\beta$	$\boldsymbol{a}^{\mathsf{T}}\mathrm{Re}\{\beta\}$	$-\frac{1}{2}\boldsymbol{a}^{\mathsf{T}}\beta^*$
$\alpha\boldsymbol{q}^{\mathsf{H}}\boldsymbol{b}$	$-\frac{1}{2}\alpha\boldsymbol{b}^{\mathsf{H}}$	$\alpha\mathrm{Re}\left\{\boldsymbol{b}^{\mathsf{T}}\right\}$
$\boldsymbol{A}\boldsymbol{q}\beta$	$\boldsymbol{A}\mathrm{Re}\{\beta\}$	$-\frac{1}{2}\boldsymbol{A}\beta^*$
$\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}$	$\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}-\frac{1}{2}(\boldsymbol{A}\boldsymbol{q})^{\mathsf{H}}$	$-\frac{1}{2}\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}+\mathrm{Re}\left\{(\boldsymbol{A}\boldsymbol{q})^{\mathsf{T}}\right\}$

Upon using the first, the second and the fourth rows of TABLE *I*, we take the gradient of $f(\boldsymbol{q})$ with respect to $\boldsymbol{q}^*$ to gield

$$\begin{aligned}
\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q})\triangleq&\left(\frac{\partial f}{\partial\boldsymbol{q}^*}\right)^{\mathsf{T}}\overset{(13)}{=}\left(\frac{\partial f}{\partial\boldsymbol{q}}\right)^{\mathsf{H}}\\
=&\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}+\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{b}-\boldsymbol{A}^{\mathsf{H}}\boldsymbol{b}\\
=&\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b}).
\end{aligned}\tag{63}$$

Then $\forall\boldsymbol{p},\boldsymbol{q}\in\mathbb{H}^n$, we obtain

$$\begin{aligned}
&f(\boldsymbol{q})-f(\boldsymbol{p})-4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\}\\
=&(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})-(\boldsymbol{A}\boldsymbol{p}-\boldsymbol{b})^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{p}-\boldsymbol{b})\\
&-2\mathrm{Re}\left\{\left(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{p}-\boldsymbol{b})\right)^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\}\\
=&\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}+\boldsymbol{p}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{p}-\boldsymbol{p}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{q}-\boldsymbol{q}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{p}\\
=&(\boldsymbol{q}-\boldsymbol{p})^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}(\boldsymbol{q}-\boldsymbol{p})\\
=&\|\boldsymbol{A}(\boldsymbol{q}-\boldsymbol{p})\|_2^2\\
\geqslant&0.
\end{aligned}\tag{64}$$

Therefore, from Theorem 3.3 we know that $f(\boldsymbol{q})$ is convex.

(**Gradient monotonicity criterion**) $\forall\boldsymbol{p},\boldsymbol{q}\in\mathbb{H}^n$, we have

$$\begin{aligned}
&\mathrm{Re}\left\{(\nabla_{\boldsymbol{p}^*}f(\boldsymbol{p})-\nabla_{\boldsymbol{q}^*}f(\boldsymbol{q}))^{\mathsf{H}}(\boldsymbol{p}-\boldsymbol{q})\right\}\\
=&\frac{1}{2}\mathrm{Re}\left\{\left(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{p}-\boldsymbol{b})-\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})\right)^{\mathsf{H}}(\boldsymbol{p}-\boldsymbol{q})\right\}\\
=&\frac{1}{2}\mathrm{Re}\left\{(\boldsymbol{p}-\boldsymbol{q})^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}(\boldsymbol{p}-\boldsymbol{q})\right\}\\
=&\frac{1}{2}(\boldsymbol{p}-\boldsymbol{q})^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}(\boldsymbol{p}-\boldsymbol{q})\\
=&\frac{1}{2}\|\boldsymbol{A}(\boldsymbol{q}-\boldsymbol{p})\|_2^2\\
\geqslant&0.
\end{aligned}\tag{65}$$

Therefore, from Theorem 3.4 we know that $f(\boldsymbol{q})$ is convex.

(**Second-order characterization criterion**) Using the third row of TABLE I, we get

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} &\triangleq \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{\partial f}{\partial \boldsymbol{q}^*}\right)^{\mathsf{T}} = \frac{\partial \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q})}{\partial \boldsymbol{q}} \\ &\overset{(63)}{=} \frac{\partial\left(\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})\right)}{\partial \boldsymbol{q}} = \frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A},\end{aligned} \tag{66}$$

and for any $\nu \in \{i, j, k\}$,

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{q}^\nu\boldsymbol{q}^*} &\triangleq \frac{\partial}{\partial \boldsymbol{q}^\nu}\left(\frac{\partial f}{\partial \boldsymbol{q}^*}\right)^{\mathsf{T}} = \frac{\partial \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q})}{\partial \boldsymbol{q}^\nu} \\ &\overset{(63)}{=} \frac{\partial\left(\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{q}-\boldsymbol{b})\right)}{\partial \boldsymbol{q}^\nu} \overset{(9)}{=} \frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{q}^\nu} = \boldsymbol{O}.\end{aligned} \tag{67}$$

Then $\forall \boldsymbol{x} \in \mathbb{H}^n, \boldsymbol{x} \neq \boldsymbol{0}$, it follows that

$$\begin{aligned}&\sum_{\nu\in\{1,i,j,k\}} \mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{H}_{\boldsymbol{q}^\nu\boldsymbol{q}^*}\boldsymbol{x}^\nu\right\} \\ =&\frac{1}{2}\mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{x}\right\} = \frac{1}{2}\boldsymbol{x}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{x} = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}\|_2^2 \geqslant 0.\end{aligned} \tag{68}$$

Therefore, by Corollary 3.1, we know that $f(\boldsymbol{q})$ is convex. ■

IV. Strongly Convex Quaternion Function: Definition and Discriminant Theorems

We shall now discuss the discriminant criteria for strongly convex quaternion functions, building upon the theorems for convexity. These criteria will be useful in designing optimization algorithms.

Definition 4.1 (Strongly convex function): The quaternion function $f(\boldsymbol{q}) : \mathcal{C} \subset \mathbb{H}^n \to \mathbb{R}$ is called strongly convex, if $\exists \sigma > 0$, $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$, $\forall \theta \in (0, 1)$,

$$f\big(\theta\boldsymbol{p}+(1-\theta)\boldsymbol{q}\big) \leqslant \theta f(\boldsymbol{p})+(1-\theta)f(\boldsymbol{q})-\frac{\sigma}{2}\theta(1-\theta)\|\boldsymbol{p}-\boldsymbol{q}\|_2^2, \tag{69}$$

where σ is the strongly convex parameter. For convenience, $f(\boldsymbol{q})$ is also called σ-strongly convex.

Based on the definition of strongly convex functions, we obtain the following equivalence theorem.

Theorem 4.1: The quaternion function $f(\boldsymbol{q}) : \mathcal{C} \subset \mathbb{H}^n \to \mathbb{R}$ is σ-strongly convex, iff $\exists \sigma > 0$, s.t. the function

$$g(\boldsymbol{q}) \triangleq f(\boldsymbol{q}) - \frac{\sigma}{2}\|\boldsymbol{q}\|_2^2 \tag{70}$$

is convex.

Proof: This is straightforward to prove, by applying Definition 3.2 and Definition 4.1. ■

Similar to convex quaternion functions, strongly convex quaternion functions also have first-order characterization, gradient monotonicity, and second-order characterization.

Theorem 4.2 (First-order characterization): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, $f(\boldsymbol{q})$ is σ-strongly convex iff $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$f(\boldsymbol{q}) \geqslant f(\boldsymbol{p}) + 4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\} + \frac{\sigma}{2}\|\boldsymbol{q}-\boldsymbol{p}\|_2^2, \tag{71}$$

where $\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})$ is defined in (15).

Proof: From Theorem 4.1, $f(\boldsymbol{p})$ is strongly convex iff $g(\boldsymbol{p}) = f(\boldsymbol{p}) - \frac{1}{2}\sigma\|\boldsymbol{p}\|_2^2$ is convex. Then, upon applying Theorem 3.3, $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$g(\boldsymbol{q}) \geqslant g(\boldsymbol{p}) + 4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*} g(\boldsymbol{p})^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\}. \tag{72}$$

Using the fourth row of TABLE I, $\nabla_{\boldsymbol{p}^*} g(\boldsymbol{p}) = \nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \frac{1}{4}\sigma\boldsymbol{p}$. Then, $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$\begin{aligned}&f(\boldsymbol{q}) - \frac{\sigma}{2}\|\boldsymbol{q}\|_2^2 \\ \geqslant& f(\boldsymbol{p}) - \frac{\sigma}{2}\|\boldsymbol{p}\|_2^2 + 4\mathrm{Re}\left\{\big(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \frac{\sigma}{4}\boldsymbol{p}\big)^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\}.\end{aligned} \tag{73}$$

Since

$$\begin{aligned}&\frac{\sigma}{2}\|\boldsymbol{q}\|_2^2 - \frac{\sigma}{2}\|\boldsymbol{p}\|_2^2 + 4\mathrm{Re}\left\{\big(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \frac{\sigma}{4}\boldsymbol{p}\big)^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\} \\ =&4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\} - \sigma\mathrm{Re}\left\{\boldsymbol{p}^{\mathsf{H}}\boldsymbol{q}\right\} + \sigma\|\boldsymbol{p}\|_2^2 \\ &+\frac{\sigma}{2}\|\boldsymbol{q}\|_2^2 - \frac{\sigma}{2}\|\boldsymbol{p}\|_2^2 \\ \overset{(32)}{=}&4\mathrm{Re}\left\{\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})^{\mathsf{H}}(\boldsymbol{q}-\boldsymbol{p})\right\} + \frac{\sigma}{2}\|\boldsymbol{q}-\boldsymbol{p}\|_2^2.\end{aligned} \tag{74}$$

Upon substituting (74) into (73), the proof follows. ■

Theorem 4.3 (Gradient monotonicity): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, $f(\boldsymbol{q})$ is σ-strongly convex iff $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$\mathrm{Re}\left\{\big(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q})\big)^{\mathsf{H}}(\boldsymbol{p}-\boldsymbol{q})\right\} \geqslant \frac{\sigma}{4}\|\boldsymbol{p}-\boldsymbol{q}\|_2^2, \tag{75}$$

where $\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p})$ is defined in (15).

Proof: From the Theorem 4.1, $f(\boldsymbol{q})$ is strongly convex iff $g(\boldsymbol{q}) = f(\boldsymbol{q}) - \frac{1}{2}\sigma\|\boldsymbol{q}\|_2^2$ is convex. Then, after applying Theorem 3.4, we have

$$\mathrm{Re}\left\{(\nabla_{\boldsymbol{p}^*} g(\boldsymbol{p}) - \nabla_{\boldsymbol{q}^*} g(\boldsymbol{q}))^{\mathsf{H}}(\boldsymbol{p}-\boldsymbol{q})\right\} \geqslant 0,\ \forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}. \tag{76}$$

Using the fourth row of TABLE I, we have $\nabla_{\boldsymbol{q}^*} g(\boldsymbol{q}) = \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q}) - \frac{1}{4}\sigma\boldsymbol{q}$, then $\forall \boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$,

$$\mathrm{Re}\left\{\big(\nabla_{\boldsymbol{p}^*} f(\boldsymbol{p}) - \frac{\sigma}{4}\boldsymbol{p} - \nabla_{\boldsymbol{q}^*} f(\boldsymbol{q}) + \frac{\sigma}{4}\boldsymbol{q}\big)^{\mathsf{H}}(\boldsymbol{p}-\boldsymbol{q})\right\} \geqslant 0. \tag{77}$$

Upon rearranging the terms in (77), we obtain (75). ■

Theorem 4.4 (Second-order characterization): Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, $f(\boldsymbol{q})$ is σ-strongly convex, iff

$$\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \frac{\sigma}{4}\boldsymbol{I}_{4n}, \tag{78}$$

where $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$ is defined in (25).

Proof: Define $g(\boldsymbol{q}) \triangleq f(\boldsymbol{q}) - \frac{1}{2}\sigma\|\boldsymbol{q}\|_2^2$, $h(\boldsymbol{q}) \triangleq 2\|\boldsymbol{q}\|_2^2 = 2\boldsymbol{q}^{\mathsf{H}}\boldsymbol{q} = 2\big(\boldsymbol{q}^{\mathsf{H}}\boldsymbol{q}\big)^\mu \overset{(6)}{=} 2\boldsymbol{q}^{\mu\mathsf{H}}\boldsymbol{q}^\mu$, $\mu \in \{1, i, j, k\}$. Upon applying the fourth row of TABLE I, we have

$$\left(\frac{\partial h}{\partial \boldsymbol{q}^{\mu*}}\right)^{\mathsf{T}} = \boldsymbol{q}^\mu, \quad \mu \in \{1, i, j, k\}. \tag{79}$$

Then, $\forall \mu \in \{1, i, j, k\}$,

$$\frac{\partial}{\partial \boldsymbol{q}^\mu}\left(\frac{\partial h}{\partial \boldsymbol{q}^{\mu*}}\right)^{\mathsf{T}} \overset{(79)}{=} \frac{\partial \boldsymbol{q}^\mu}{\partial \boldsymbol{q}^\mu} = \boldsymbol{I}_n, \tag{80}$$

and $\forall \mu, \nu \in \{1, i, j, k\}$, $\mu \neq \nu$,

$$\frac{\partial}{\partial \boldsymbol{q}^{\nu}} \left(\frac{\partial h}{\partial \boldsymbol{q}^{\mu *}} \right)^{\mathsf{T}} \overset{(79)}{=} \frac{\partial \boldsymbol{q}^{\mu}}{\partial \boldsymbol{q}^{\nu}} = \boldsymbol{O}. \tag{81}$$

By (25), the augmented quaternion Hessian matrix of h is $\boldsymbol{I}_{4n}$. Therefore, the augmented quaternion Hessian matrix of g is $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} - \frac{1}{4}\sigma \boldsymbol{I}_{4n}$. From Theorem 4.1, $f(\boldsymbol{q})$ is strongly convex iff $g(\boldsymbol{q})$ is convex. Then upon applying Theorem 3.7, $g(\boldsymbol{q}) = f(\boldsymbol{q}) - \frac{1}{2}\sigma \|\boldsymbol{q}\|_2^2$ is convex iff $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} - \frac{1}{4}\sigma \boldsymbol{I}_{4n} \succeq \boldsymbol{O}$. ■

Corollary 4.1: Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. Then, the following three propositions are equivalent:

(a) $f(\boldsymbol{q})$ is σ-strongly convex;

(b) $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \frac{1}{4}\sigma \boldsymbol{I}_{4n}$;

(c) $\sum\limits_{\nu \in \{1,i,j,k\}} \mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}} \boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^*} \boldsymbol{x}^{\nu}\right\} - \frac{1}{4}\sigma \|\boldsymbol{x}\|_2^2 \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{H}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$.

Proof: According to Theorem 4.4, (a) is equivalent to (b), so we only need to prove that (b) is equivalent to (c). By Corollary 2.1, $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*}$ is a Hermite matrix, so $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} - \frac{1}{4}\sigma \boldsymbol{I}_{4n}$ is also Hermite matrix. Then, $\forall \boldsymbol{x}_{\mathcal{H}} \in \mathcal{H}$, $\boldsymbol{x}_{\mathcal{H}} \neq \boldsymbol{0}$, we have

$$\begin{aligned} & \boldsymbol{x}_{\mathcal{H}}^{\mathsf{H}} \left(\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} - \frac{\sigma}{4} \boldsymbol{I}_{4n} \right) \boldsymbol{x}_{\mathcal{H}} \\ &= \boldsymbol{x}_{\mathcal{H}}^{\mathsf{H}} \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \boldsymbol{x}_{\mathcal{H}} - \frac{\sigma}{4} \boldsymbol{x}_{\mathcal{H}}^{\mathsf{H}} \boldsymbol{x}_{\mathcal{H}} \\ &\overset{(31)}{=} \sum_{\mu,\nu \in \{1,i,j,k\}} \boldsymbol{x}^{\mu \mathsf{H}} \boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^{\mu *}} \boldsymbol{x}^{\nu} - \sigma \boldsymbol{x}^{\mathsf{H}} \boldsymbol{x} \\ &\overset{(20)}{=} 4 \sum_{\nu \in \{1,i,j,k\}} \mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}} \boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^*} \boldsymbol{x}^{\nu}\right\} - \sigma \|\boldsymbol{x}\|_2^2. \end{aligned} \tag{82}$$

Therefore, $\forall \boldsymbol{x} \in \mathbb{H}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$,

$$\begin{aligned} & \sum_{\nu \in \{1,i,j,k\}} \mathrm{Re}\left\{\boldsymbol{x}^{\mathsf{H}} \boldsymbol{H}_{\boldsymbol{q}^{\nu}\boldsymbol{q}^*} \boldsymbol{x}^{\nu}\right\} - \frac{\sigma}{4} \|\boldsymbol{x}\|_2^2 \geqslant 0, \\ \Leftrightarrow \quad & \boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} - \frac{\sigma}{4} \boldsymbol{I}_{4n} \succeq \boldsymbol{O}. \end{aligned} \tag{83}$$

This completes the proof. ■

Upon applying Lemma 3.1, we can obtain a necessary condition for σ-strongly convex quaternion functions.

Theorem 4.5: Consider a convex set $\mathcal{C} \subset \mathbb{H}^n$ and a second-order continuous real-differentiable quaternion function $f(\boldsymbol{q}) : \mathcal{C} \to \mathbb{R}$. If $f(\boldsymbol{q})$ is σ-strongly convex, then

$$\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} \succeq \frac{\sigma}{4} \boldsymbol{I}_n, \tag{84}$$

where $\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*}$ is the quaternion Hessian matrix, defined in (17).

Proof: Note that $f(\boldsymbol{q})$ is σ-strongly convex, and upon applying Theorem 4.4, we have $\boldsymbol{H}_{\mathcal{H}\mathcal{H}^*} \succeq \frac{1}{4}\sigma \boldsymbol{I}_{4n}$. By (25) and Lemma 3.1, we finally obtain $\boldsymbol{H}_{\boldsymbol{q}\boldsymbol{q}^*} \succeq \frac{1}{4}\sigma \boldsymbol{I}_n$. ■

V. Convex Quaternion Optimization Problems and Their Applications in Signal Processing

We now proceed to introduce the convex quaternion problem and its fundamental theorem. This is followed by several applications of convex quaternion optimization in communications, highlighting its practical significance.

A. Convex Quaternion Optimization Problems

Similar to convex real and complex optimization problems, convex quaternion optimization problems generally have a structure which consist of the minimization of a convex quaternion function subject to (shortened to s.t.) quaternion affine equality constraints and inequality constraints defined by convex quaternion functions, as follows

$$\begin{aligned} \min_{\boldsymbol{q} \in \mathbb{H}^n} \; & f_0(\boldsymbol{q}) \\ \text{s.t.} \; & \boldsymbol{A}\boldsymbol{q} = \boldsymbol{b}, \\ & f_i(\boldsymbol{q}) \leqslant 0, \; i = 1, \ldots, P \end{aligned} \tag{85}$$

where $f_i : \mathbb{H}^n \to \mathbb{R}$, $i = 0, 1, \ldots, P$ is convex, $\boldsymbol{A} \in \mathbb{H}^{m \times n}$, $\boldsymbol{b} \in \mathbb{H}^m$. The problem field is $\mathcal{F} \triangleq \bigcap_{i=0}^{P} \boldsymbol{dom} f_i$, and feasible set is $\mathcal{C} \triangleq \{\boldsymbol{q} \in \mathcal{F} \mid f_i(\boldsymbol{q}) \leqslant 0, \; i = 1, \ldots, P, \; \boldsymbol{A}\boldsymbol{q} = \boldsymbol{b}\}$. From Definition 3.2, Example 3.1 and Example 3.2, the sets $\mathcal{D} \triangleq \{\boldsymbol{q} \in \mathbb{H}^n \mid \boldsymbol{A}\boldsymbol{q} = \boldsymbol{b}\}$, $\mathcal{E}_i \triangleq \{\boldsymbol{q} \in \mathbb{H}^n \mid f_i(\boldsymbol{q}) \leqslant 0\}$, $i = 1, \ldots, P$ and $\boldsymbol{dom} f_i$, $i = 0, 1, \ldots, P$ are convex. Therefore, the set $\mathcal{C} = \mathcal{D} \bigcap \left(\bigcap_{i=1}^{P} \mathcal{E}_i \right) \bigcap \mathcal{F}$ is also convex.

When studying the convexity of quaternion functions, we utilized the augmented quaternion vectors and augmented real vectors. Similarly, when studying the properties of quaternion convex optimization problems, we also need to utilize the augmented quaternion and the augmented real convex optimization settings.

The convex augmented quaternion optimization problem of (85) is given by [22]

$$\begin{aligned} \min_{\boldsymbol{q}_{\mathcal{H}} \in \mathcal{H}} \; & f_0(\boldsymbol{q}_{\mathcal{H}}) \\ \text{s.t.} \; & \boldsymbol{A}_{\mathcal{H}} \boldsymbol{q}_{\mathcal{H}} = \boldsymbol{b}_{\mathcal{H}}, \\ & f_i(\boldsymbol{q}_{\mathcal{H}}) \leqslant 0, \; i = 1, \ldots, P \end{aligned} \tag{86}$$

where $f_i : \mathcal{H} \to \mathbb{R}$, $i = 0, 1, \ldots, P$ is convex, $\boldsymbol{A}_{\mathcal{H}} \triangleq \mathrm{diag}\left(\boldsymbol{A}, \boldsymbol{A}^i, \boldsymbol{A}^j, \boldsymbol{A}^k\right) \in \mathbb{H}^{4m \times 4n}$, $\boldsymbol{A} \in \mathbb{H}^{m \times n}$, and $\boldsymbol{b}_{\mathcal{H}} \triangleq \left(\boldsymbol{b}^{\mathsf{T}}, \boldsymbol{b}^{i\mathsf{T}}, \boldsymbol{b}^{j\mathsf{T}}, \boldsymbol{b}^{k\mathsf{T}}\right)^{\mathsf{T}} \in \mathbb{H}^{4m}$.

The convex augmented real convex optimization problem of (85) is given by [22]

$$\begin{aligned} \min_{\boldsymbol{q}_{\mathcal{R}} \in \mathcal{R}} \; & f_0(\boldsymbol{q}_{\mathcal{R}}) \\ \text{s.t.} \; & \boldsymbol{A}_{\mathcal{R}} \boldsymbol{q}_{\mathcal{R}} = \boldsymbol{b}_{\mathcal{R}}, \\ & f_i(\boldsymbol{q}_{\mathcal{R}}) \leqslant 0, \; i = 1, \ldots, P \end{aligned} \tag{87}$$

where $f_i : \mathcal{R} \to \mathbb{R}$, $i = 0, 1, \ldots, P$ is convex, together with $\boldsymbol{A}_{\mathcal{R}} \triangleq \frac{1}{4} \boldsymbol{J}_m^{\mathsf{H}} \boldsymbol{A}_{\mathcal{H}} \boldsymbol{J}_n \in \mathbb{R}^{4m \times 4n}$, $\boldsymbol{A}_{\mathcal{H}} \in \mathbb{H}^{4m \times 4n}$, and $\boldsymbol{b}_{\mathcal{R}} \triangleq \frac{1}{4} \boldsymbol{J}_m^{\mathsf{H}} \boldsymbol{b}_{\mathcal{H}} \in \mathbb{R}^{4m}$.

Lemma 5.1 ([22]): The convex quaternion optimization problem in (85), the convex augmented quaternion optimization problem in (86), and the convex augmented real optimization problem in (87) are equivalent.

Theorem 5.1: For the convex quaternion optimization problem in (85), any local optimal solution is also the global optimal solution.

Proof: We already know [43] that for the real convex optimization problem in (87), any local optimal solution, for example $\bar{\boldsymbol{q}}_{\mathcal{R}}$, is also the global optimal solution. Then, from Lemma 5.1, the local optimal solution $\bar{\boldsymbol{q}}_{\mathcal{H}} \triangleq \left(\bar{\boldsymbol{q}}^{\mathsf{T}}, \bar{\boldsymbol{q}}^{i\mathsf{T}}, \bar{\boldsymbol{q}}^{j\mathsf{T}}, \bar{\boldsymbol{q}}^{k\mathsf{T}}\right)^{\mathsf{T}} = \boldsymbol{J}_n \bar{\boldsymbol{q}}_{\mathcal{R}}$ is also global, in the augmented convex quaternion optimization problem in (86). Therefore, the local optimal solution, $\bar{\boldsymbol{q}}$, is also global, in the convex quaternion optimization problem in (85). ■

B. Applications of Convex Quaternion Optimization in Signal Processing

Application 5.1 (Quaternion linear mean-square error filter): The quaternion minimum mean-square error (MSE) filter can be specified as

$$\min_{\boldsymbol{w}\in\mathbb{H}^n} \mathcal{J}(\boldsymbol{w}) \triangleq E\left\{|e(n)|^2\right\} = E\left\{|d(n)-y(n)|^2\right\}, \quad (88)$$

where $y(n)=\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}(n)$, $\boldsymbol{x}(n)\in\mathbb{H}^n$ is the input vector, $\boldsymbol{w}\in\mathbb{H}^n$ is the filter weight vector, and $d(n)\in\mathbb{H}$ is the desired sequence. By the definition of the modulus, we have

$$\begin{aligned}\mathcal{J}(\boldsymbol{w}) =&E\left\{|d(n)-\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}(n)|^2\right\}\\ =&E\left\{\left(d(n)-\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}(n)\right)\left(d(n)-\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}(n)\right)^*\right\}\\ =&\boldsymbol{w}^{\mathsf{H}}E\{\boldsymbol{x}(n)\boldsymbol{x}^{\mathsf{H}}(n)\}\boldsymbol{w}-E\{d(n)\boldsymbol{x}^{\mathsf{H}}(n)\}\boldsymbol{w}\\ &-\boldsymbol{w}^{\mathsf{H}}E\{\boldsymbol{x}(n)d^*(n)\}+E\{d(n)d^*(n)\}\\ =&\boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w}-\boldsymbol{p}^{\mathsf{H}}\boldsymbol{w}-\boldsymbol{w}^{\mathsf{H}}\boldsymbol{p}+\sigma_d^2,\end{aligned} \quad (89)$$

where $\boldsymbol{R}=E\{\boldsymbol{x}(n)\boldsymbol{x}^{\mathsf{H}}(n)\}$ denotes the quaternion-valued input correlation matrix, $\boldsymbol{p}=E\{\boldsymbol{x}(n)d^*(n)\}$ is the crosscorrelation vector between the desired response and the input signal, $\sigma_d^2=E\{d(n)d^*(n)\}$ is the power of the desired response. By (62) in Example 3.3, we know that $\mathcal{J}(\boldsymbol{w})$ is convex. Similarly to (63), we take the gradient of $\mathcal{J}(\boldsymbol{w})$ with respect to $\boldsymbol{w}^*$, and set the result to $\mathbf{0}$ to obtain

$$\nabla_{\boldsymbol{w}^*}\mathcal{J}(\boldsymbol{w})=\left(\frac{\partial\mathcal{J}}{\partial\boldsymbol{w}}\right)^{\mathsf{H}}=\frac{1}{2}\boldsymbol{R}\boldsymbol{w}-\frac{1}{2}\boldsymbol{p}=\mathbf{0}. \quad (90)$$

Using (90) and Theorem 5.1, we arrive at gives the closed-form optimal solution

$$\bar{\boldsymbol{w}}=\boldsymbol{R}^{-1}\boldsymbol{p}. \quad (91)$$

Application 5.2 (Quaternion projection on affine equality constraint): The quaternion projection problem can be described as

$$\begin{aligned}&\min_{\boldsymbol{x}\in\mathbb{H}^n} \|\boldsymbol{x}-\boldsymbol{y}\|_2^2\\ &\text{s.t. } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{aligned} \quad (92)$$

where $\boldsymbol{y}\in\mathbb{H}^n$, $\boldsymbol{b}\in\mathbb{H}^p$, $\boldsymbol{A}\in\mathbb{H}^{p\times n}$ and $\operatorname{rank}(\boldsymbol{A})=p<n$. Applying Example 3.3, $f(\boldsymbol{x})=\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$ is convex, and $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ is an affine equality constraint. Therefore, the quaternion optimization problem in (92) is convex.

Using the methed of Lagrange multipliers [22, 36], we have

$$\begin{aligned}&\mathcal{L}(\boldsymbol{x},\,\boldsymbol{\lambda})\\ =&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+\mathrm{Re}\left\{\boldsymbol{\lambda}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\right\}\\ =&(\boldsymbol{x}-\boldsymbol{y})^{\mathsf{H}}(\boldsymbol{x}-\boldsymbol{y})+\frac{1}{2}\boldsymbol{\lambda}^{\mathsf{H}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+\frac{1}{2}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathsf{H}}\boldsymbol{\lambda}\\ =&\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}+\left(\frac{1}{2}\boldsymbol{\lambda}^{\mathsf{H}}\boldsymbol{A}-\boldsymbol{y}^{\mathsf{H}}\right)\boldsymbol{x}+\boldsymbol{x}^{\mathsf{H}}\left(\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{\lambda}-\boldsymbol{y}\right)\\ &+\boldsymbol{y}^{\mathsf{H}}\boldsymbol{y}-\frac{1}{2}\boldsymbol{\lambda}^{\mathsf{H}}\boldsymbol{b}-\frac{1}{2}\boldsymbol{b}^{\mathsf{H}}\boldsymbol{\lambda},\end{aligned} \quad (93)$$

where $\boldsymbol{\lambda}\in\mathbb{H}^p$ denotes the set of Lagrange multipliers. Finding the gradient of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})$ with respect to $\boldsymbol{x}^*$ in the same way as in (63) and setting the result to $\mathbf{0}$, we have

$$\begin{aligned}\nabla_{\boldsymbol{x}^*}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})=&\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{x}}\right)^{\mathsf{H}}\\ =&\frac{1}{2}\boldsymbol{x}+\left(\frac{1}{2}\boldsymbol{\lambda}^{\mathsf{H}}\boldsymbol{A}-\boldsymbol{y}^{\mathsf{H}}\right)^{\mathsf{H}}-\frac{1}{2}\left(\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{\lambda}-\boldsymbol{y}\right)\\ =&\frac{1}{2}\boldsymbol{x}-\frac{1}{2}\boldsymbol{y}+\frac{1}{4}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{\lambda}\\ =&\mathbf{0},\end{aligned} \quad (94)$$

which leads to

$$\boldsymbol{x}=\boldsymbol{y}-\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{\lambda}. \quad (95)$$

A combinition of (95) with the constraint $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ yields

$$\begin{aligned}&\boldsymbol{A}\left(\boldsymbol{y}-\frac{1}{2}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{\lambda}\right)=\boldsymbol{b}\\ \Rightarrow\quad&\boldsymbol{\lambda}=2\left(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}\right)^{-1}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}).\end{aligned} \quad (96)$$

Substituting (96) into (95) and applying Theorem 5.1, we obtain the following optimal solution

$$\bar{\boldsymbol{x}}=\boldsymbol{y}+\boldsymbol{A}^{\mathsf{H}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}\right)^{-1}(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{y}). \quad (97)$$

Application 5.3 (Quaternion minimum variance beamforming): The problem of quaternion variance beamforming minimization can be described as

$$\begin{aligned}&\min_{\boldsymbol{w}\in\mathbb{H}^n} f(\boldsymbol{w})\triangleq\boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w}\\ &\text{s.t. } \boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}=1,\end{aligned} \quad (98)$$

where $\boldsymbol{w}\in\mathbb{H}^n$ is the beamformer weight vector, $\boldsymbol{a}\in\mathbb{H}^n$ is the steering vector, and $\boldsymbol{R}^{\mathsf{H}}=\boldsymbol{R}\in\mathbb{H}^{n\times n}$ is positive definite.

We will next prove that the problem in (98) is a convex quaternion optimization problem. Using the fourth row of TABLE I, we have

$$\nabla_{\boldsymbol{w}^*}f(\boldsymbol{w})=\left(\frac{\partial f}{\partial\boldsymbol{w}}\right)^{\mathsf{H}}=\left(\boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}-\frac{1}{2}(\boldsymbol{R}\boldsymbol{w})^{\mathsf{H}}\right)^{\mathsf{H}}=\frac{1}{2}\boldsymbol{R}\boldsymbol{w}. \quad (99)$$

Then, $\forall\boldsymbol{v},\boldsymbol{w}\in\mathbb{H}^n$,

$$\begin{aligned}&\mathrm{Re}\left\{\left(\nabla_{\boldsymbol{v}^*}f(\boldsymbol{v})-\nabla_{\boldsymbol{w}^*}f(\boldsymbol{w})\right)^{\mathsf{H}}(\boldsymbol{v}-\boldsymbol{w})\right\}\\ =&\frac{1}{2}\mathrm{Re}\left\{(\boldsymbol{R}\boldsymbol{v}-\boldsymbol{R}\boldsymbol{w})^{\mathsf{H}}(\boldsymbol{v}-\boldsymbol{w})\right\}\\ =&\frac{1}{2}(\boldsymbol{v}-\boldsymbol{w})^{\mathsf{H}}\boldsymbol{R}(\boldsymbol{v}-\boldsymbol{w})\\ \geqslant&0.\end{aligned} \quad (100)$$

From Theorem 3.4, it follows that $f(\boldsymbol{w})$ is convex, and $\boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}=1$ is an affine equality constraint. Therefore, the problem in (98) is a convex quaternion optimization problem.

The Lagrangian of problem in (98) is given by [22, 36]

$$\mathcal{L}(\boldsymbol{w},\lambda)=\boldsymbol{w}^{\mathsf{H}}\boldsymbol{R}\boldsymbol{w}+\lambda\left(\boldsymbol{w}^{\mathsf{H}}\boldsymbol{a}-1\right),\quad\lambda\in\mathbb{R}, \quad (101)$$

which is a real-valued function of $\boldsymbol{w} \in \mathbb{H}^n$. Using the second and the fourth rows of TABLE I, and setting $\nabla_{\boldsymbol{w}^*}\mathcal{L}(\boldsymbol{w},\lambda) = \mathbf{0}$, we have

$$\begin{aligned}\nabla_{\boldsymbol{w}^*}\mathcal{L}(\boldsymbol{w},\lambda) = \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}\right)^{\mathsf{H}} = \frac{1}{2}\boldsymbol{R}\boldsymbol{w} - \frac{1}{2}\lambda\boldsymbol{a} = \mathbf{0} \\ \Rightarrow \quad \boldsymbol{w} = \lambda \boldsymbol{R}^{-1}\boldsymbol{a}.\end{aligned} \tag{102}$$

Upon substituting (102) into $\boldsymbol{a}^{\mathsf{H}}\boldsymbol{w} = 1$, we obtain

$$\lambda \boldsymbol{a}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{a} = 1 \quad \Rightarrow \quad \lambda = \frac{1}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{a}}. \tag{103}$$

Therefore, upon applying Theorem 5.1, the closed-form optimal solution is obtained as

$$\bar{\boldsymbol{w}} = \frac{\boldsymbol{R}^{-1}\boldsymbol{a}}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{a}}. \tag{104}$$

VI. Conclusions

We have established the theory of convex quaternion optimization based on the GHR calculus, which is an enabling methodology in the field of quaternion optimization and its applications in quaternion signal processing and machine learning. Our study has resulted in the development of five discriminant theorems for convex functions in the quaternion field, utilizing (20), (23), (25), (26), and (27). Furthermore, we have provided the definition and four discriminant criteria for strongly convex functions by employing the results for convex quaternion functions. In addition, we have presented a fundamental theorem for the optimality of convex quaternion optimization problems and three applications in signal processing, which have both enriched the theory of convex quaternion optimization and provided a theoretical foundation for quaternion signal processing. However, the convexity of non-differentiable quaternion functions by the GHR calculus still remains an open area, and this work provides a foundation and an avenue for further research in this direction.

References

[1] W. R. Hamilton, "On a new species of imaginary quantities, connected with the theory of quaternions," in *Proceedings of the Royal Irish Academy (1836-1869)*, vol. 2. JSTOR, 1840, pp. 424–434.

[2] L. Qi, Z. Luo, Q. Wang, and X. Zhang, "Quaternion matrix optimization: Motivation and analysis," *Journal of Optimization Theory and Applications*, vol. 193, no. 1-3, pp. 621–648, 2022.

[3] Z. Jia, Q. Jin, M. K. Ng, and X. Zhao, "Non-local robust quaternion matrix completion for large-scale color image and video inpainting," *IEEE Transactions on Image Processing*, vol. 31, pp. 3868–3883, 2022.

[4] C. C. Took and D. P. Mandic, "Augmented second-order statistics of quaternion random signals," *Signal Processing*, vol. 91, no. 2, pp. 214–224, 2011.

[5] J. Flamant, S. Miron, and D. Brie, "Quaternion non-negative matrix factorization: Definition, uniqueness, and algorithm," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1870–1883, 2020.

[6] C. C. Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4427–4431, 2010.

[7] H. Zhang, Z. Wang, D. Chen, S. Zhu, and D. Xu, "Quaternion extreme learning machine based on real augmented representation," *IEEE Signal Processing Letters*, vol. 30, pp. 175–179, 2023.

[8] S. Walia, K. Kumar, and M. Kumar, "Unveiling digital image forgeries using Markov based quaternions in frequency domain and fusion of machine learning algorithms," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 4517–4532, 2023.

[9] B. C. Ujang, C. C. Took, and D. P. Mandic, "Quaternion-valued nonlinear adaptive filtering," *IEEE Transactions on Neural Networks*, vol. 22, no. 8, pp. 1193–1206, 2011.

[10] J. Flamant, N. Le Bihan, and P. Chainais, "Time-frequency analysis of bivariate signals," *Applied and Computational Harmonic Analysis*, vol. 46, no. 2, pp. 351–383, 2019.

[11] T. Ogunfunmi and C. Safarian, "The quaternion stochastic information gradient algorithm for nonlinear adaptive systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5909–5921, 2019.

[12] E. C. Mengüç, "Design of quaternion-valued second-order Volterra adaptive filters for nonlinear 3-D and 4-D signals," *Signal Processing*, vol. 174, p. 107619, 2020.

[13] Y. Xia, S. Tao, Z. Li, M. Xiang, W. Pei, and D. P. Mandic, "Full mean square performance bounds on quaternion estimators for improper data," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4093–4106, 2019.

[14] S. Enshaeifar, S. Kouchaki, C. Cheong Took, and S. Sanei, "Quaternion singular spectrum analysis of electroencephalogram with application in sleep analysis," *IEEE Transactions on Neural Systems and Rehabilitation*, vol. 24, no. 1, pp. 57–67, Jan. 2016.

[15] Z. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

[16] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. MIT Press, 2012.

[17] M. Jaggi, "Sparse convex optimization methods for machine learning," Ph.D. dissertation, ETH Zürich, 2011.

[18] N. Krejić, N. K. Jerinkić, and T. Ostojić, "An inexact restoration-nonsmooth algorithm with variable accuracy for stochastic nonsmooth convex optimization problems in machine learning and stochastic linear complementarity problems," *Journal of Computational and Applied Mathematics*, vol. 423, p. 114943, 2023.

[19] Y. Xia and D. P. Mandic, "Complementary mean square analysis of augmented CLMS for second-order noncircular Gaussian signals," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1413–1417, 2017.

[20] Y. Xia and D. P. Mandic, "A full mean square analysis of CLMS for second-order noncircular inputs," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5578–5590, 2017.

[21] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi,

M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, 2010.

[22] J. Flamant, S. Miron, and D. Brie, "A general framework for constrained convex quaternion optimization," *IEEE Transactions on Signal Processing*, vol. 70, pp. 254–267, 2022.

[23] Y. Liu, Y. Zheng, J. Lu, J. Cao, and L. Rutkowski, "Constrained quaternion-variable convex optimization: A quaternion-valued recurrent neural network approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 1022–1035, 2019.

[24] D. Xu, C. Jahanchahi, C. C. Took, and D. P. Mandic, "Enabling quaternion derivatives: The generalized HR calculus," *Royal Society Open Science*, vol. 2, no. 8, p. 150255, 2015.

[25] A. Hjørungnes, *Complex-valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, 2011.

[26] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.

[27] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," in *IEE Proceedings H (Microwaves, Optics and Antennas)*, vol. 130, no. 1. IET Digital Library, 1983, pp. 11–16.

[28] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, "Multilayer perceptrons to approximate quaternion valued functions," *Neural Networks*, vol. 10, no. 2, pp. 335–342, 1997.

[29] M. Yoshida, Y. Kuroe, and T. Mori, "A model of hopfield-type quaternion neural networks and its energy function," in *Neural Information Processing*. Springer, 2004, pp. 110–115.

[30] E. C. Mengüç, "Novel quaternion-valued least-mean kurtosis adaptive filtering algorithm based on the GHR calculus," *IET Signal Processing*, vol. 12, no. 4, pp. 487–495, 2018.

[31] C. C. Took and Y. Xia, "Multichannel quaternion least mean square algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 8524–8527.

[32] T. Parcollet, M. Morchid, and G. Linarès, "A survey of quaternion neural networks," *Artificial Intelligence Review*, vol. 53, pp. 2957–2982, 2020.

[33] D. Xu, Y. Xia, and D. P. Mandic, "Optimization in quaternion dynamic systems: Gradient, Hessian, and learning algorithms," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 249–261, 2015.

[34] J. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*. Springer Science, 1997.

[35] T. A. Ell and S. J. Sangwine, "Quaternion involutions and anti-involutions," *Computers & Mathematics with Applications*, vol. 53, no. 1, pp. 137–143, 2007.

[36] D. Xu and D. P. Mandic, "The theory of quaternion matrix derivatives," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1543–1556, 2015.

[37] A. Sudbery, "Quaternionic analysis," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 85, no. 2. Cambridge University Press, 1979, pp. 199–225.

[38] J. Vía, D. Ramírez, and I. Santamaría, "Properness and widely linear processing of quaternion random vectors," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3502–3515, 2010.

[39] D. P. Mandic, C. Jahanchahi, and C. C. Took, "A quaternion gradient operator and its applications," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 47–50, 2011.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[41] Y. Nesterov, *Lectures on Convex Optimization*. Springer, 2018.

[42] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.

[43] C. Chi, W. Li, and C. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. CRC Press, 2017.