JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 71, 187–209 (1979)

Quaternions, Fréchet Differentiation, and Some Equations of Mathematical Physics
1. Critical Point Theory

V. KOMKOV

Mathematical Reviews, University of Michigan, Ann Arbor, Michigan 48109

Submitted by C. L. Dolph

Following an introduction discussing some properties of maps $Q \to Q$ and $Q \times Q \to Q$, where Q denotes the ring of quaternions, it is shown that many equations of mathematical physics can be written in this formalism. A concept of Fréchet differentiation is given in this setting, in a manner analogous to the usual definition. Variational principles are derived. The physical examples involve elasticity, motion of rigid bodies, fluid flow and Maxwell's equations of electromagnetic field theory.

1. AN INTRODUCTORY DISCUSSION

By a quaternion we shall understand a form

$$q = a_0(\mathbf{x})\mathbf{1} + a_1(\mathbf{x})\mathbf{i} + a_2(\mathbf{x})\mathbf{j} + a_3(\mathbf{x})\mathbf{K}, \mathbf{x} \in \mathbf{R}^n,$$

where the a_i, $i = 0, 1, 2, 3$, are functions from $\mathbf{R}^n$ into $\mathbf{C}$, satisfying the usual algebraic rules of a quaternion ring (See Appendix 1). In what follows we shall consider only the cases where $n = 0, 1, 2,$ or 3, i.e., $\mathbf{x} = \{x_0, x_1, x_2, x_3\}$ at most. Suppose that $f: Q \to Q$, and the $a_i(\mathbf{x})$, $i = 0, 1, 2, 3$, are functions which belong to the Hilbert space $L_2(\Omega)$, $\Omega \subset \mathbf{R}^n$.

We define the Gâteaux derivative of $f(q)$ (with respect to q) at the point $q = \tilde{q}$ in the direction of q_1 to be the quantity $f'_{q_1}(q)|_{q=\tilde{q}} = q_2$ such that, for any $\epsilon > 0$, $f(q \pm \epsilon q_1) - f(q) = \epsilon(q_1 \times q_2)$, i.e., a "left" linear form in q_1. (See Lemma 1 for the reason why terms in higher power of ϵ are not even mentioned.) If $q = q(\mathbf{x})$, we would define

$$\{q_0 \times q_1\}_\Omega = \int_\Omega (q_0 \times q_1)\, d\mathbf{x}$$

and find $q_2(\mathbf{x})$ by the rule

$$\epsilon \int (q_1(\mathbf{x}) \times q_2(\mathbf{x}))\, d\mathbf{x} \approx f(q(\mathbf{x}) + \epsilon q_1(\mathbf{x})) - f(q(\mathbf{x})).$$

0022-247X/79/090187–23\$02.00/0

The symbol $\approx$ implying that the remainder is of the order $o(\epsilon^2 q_1(\mathbf{x}))$ for small ϵ. As before, $\times$ denotes quaternion multiplication.

It will become apparent from physical examples that the choice of the quaternionic product has to be unique if physical laws are to be represented by Fréchet differentiation.

Gâteaux Differentiability

Let us define the Gâteaux derivative of $f(q), f: Q \to Q$, in the direction of a quaternion $q_1 \neq \varnothing$, to be q_2 such that $\epsilon(q_2 \times q_3) = f(q + \epsilon q_1) - f(q)$ for any $\epsilon > 0$.

LEMMA 1. *q_2 is uniquely defined.*

Proof. This follows immediately from the divisibility property of the ring of quaternions. For any fixed $\mathbf{x} = \bar{\mathbf{x}}$ we have a unique quaternion

$$q_2 = \mathbf{1}c_0 + \mathbf{i}c_1 + \mathbf{j}c_2 + \mathbf{K}c_3$$

such that

$$q_2 \times \epsilon q_1(\bar{\mathbf{x}}) = f(q(\bar{\mathbf{x}} + \epsilon q_1(\bar{\mathbf{x}})) = -f(q(\bar{\mathbf{x}})).$$

Defining $q_2(\bar{\mathbf{x}})$ to be that quaternion for each $\bar{\mathbf{x}}$ completes the proof of the uniqueness of the definition.

Comment. The choice of the "left" definition $\partial f(q)/\partial q = q_2$ such that $\epsilon(q_2 \times q_1) = f(q + \epsilon q_1) - f(q)$ is arbitrary. A different value would have resulted if we had defined $\partial f/\partial q = q_2$ such that $\epsilon(q_1 \times q_2) = f(q + \epsilon q_1) - f(q)$. As long as the definition remains consistent, and "right" or "left" differentiation definitions are used consistently, either one can be applied. For bilinear products, such as $Q = \{q_0, q_1\} = \int_\Omega (q_0(\mathbf{x}) \times q_1(\mathbf{x}))\, d\mathbf{x}$, the most convenient definition mixing the "left" and the "right" definitions results in $\partial Q/\partial q_0 = q_1$, $\partial Q/\partial q_1 = q_0$.

This turns out to be the definition which confirms the formal manipulation rules of mathematical physics. Of course, we can define the Fréchet derivative of any quaternion $q_0(\mathbf{x})$ with respect to any other quaternion $q_1(\mathbf{x}) \neq \varnothing$, in an identical manner.

Say $q_1 \neq \varnothing$. Thee there exists q_3 such that $q_0(q_1) = q_3 \times q_1$. We form the difference in the direction of an arbitrary quaternion q_2

$$q_3 \times (q_1 + \epsilon q_2) - q_3 \times q_1 = \epsilon q_3 \times q_2 .$$

Hence $\partial q_0/\partial q_1 = q_3$ by definition. Using the "right" instead of the "left" representation $q_0(q_1) = q_1 \times q_3$, we obtain an identical result

$$(q_1 + \epsilon q_2) \times q_3 - q_1 \times q_3 = \epsilon q_2 \times q_3 .$$

Hence $\partial q_0/\partial q_1 = q_3$. Similarly, $\partial(Aq_0)/\partial q_0 = A$ for any linear operator A, since $A(q_0 + \epsilon q_1) - Aq_1 = \epsilon Aq_1$ and $\partial(A \times q_0)/\partial q_0 = A$ if the "right" definition is adopted, while $\partial(A \times q_0)/\partial q_0 = A^*$ if the "left" definition is used.

From the above discussion the definition of second derivatives of a bilinear functional of the form $F(q_0, q_1) \to \mathbf{R}$ follows logically.

By the quaternionic version of the Lax–Milgram theorem, $F(q_0, q_1)$ must be of the form $F(q_0, q_1) = \int (Aq_0 \times q)\, d\mathbf{x} = \{Aq_0, q_1\}$, where A is a linear operator; i.e., the mapping of $q_0(\mathbf{x})$ into $Aq_0(\mathbf{x})$ is linear in Ω.

It follows from the definition that

$$\partial F(q_0, q_1)/\partial q_0 = A^* q_1,$$

$$\partial F(q_0, q_1)/\partial q_1 = Aq_0.$$

We observe that second derivatives do not obey Tonelli's law, i.e.,

$$\partial^2 F/\partial q_0\, \partial q_1 \neq \partial^2 F/\partial q_1\, \partial q_0.$$

LEMMA. *Every binary function $f: Q \times Q \to Q$ which linearly depends on a given quaternion $\hat{q}$ can be written in the form $f(q, \hat{q}) = \hat{q} \times \phi(q)$.*

Proof. The hypothesis implies that $f(q, c\hat{q}) = cf(q, \hat{q})$ for any constant c. Clearly if $\hat{q} = \varnothing$ then $f = \varnothing$. If $\hat{q} \neq \varnothing$, define $\phi(q) = \hat{q}^{-1} \times f(q, \hat{q})$ and define $\phi(q)$ arbitrarily if $\hat{q} = \varnothing$. Q.E.D.

LEMMA 2. *If f is a polynomial, then the Gâteaux derivative of f is independent of the direction, i.e., f is Fréchet differentiable, and the derivative is a scalar.*

Proof. A straightforward computation using only the definition.

Comment. This statement does not have to be true in general. Consider the following counterexample:

$$f(q) = q_+(\mathbf{x}) = a_{0_+}(\mathbf{x})\, \mathbf{1} + a_{1_+}(\mathbf{x})\, \mathbf{i} + a_{2_+}(\mathbf{x})\, \mathbf{j} + a_{3_+}(\mathbf{x})\, \mathbf{K},$$

where

$$\begin{aligned} a_{i_+}(\mathbf{x}) &= 0 \qquad \text{if} \quad a_i(\mathbf{x}) < 0 \\ &= a_i(\mathbf{x}) \qquad \text{if} \quad a_i(\mathbf{x}) \geqslant 0, \quad i = 0, 1, 2, 3. \end{aligned}$$

Trying to solve for q_2 the equation

$$f(q + \epsilon q_1) - f(q) = (q + \epsilon q_1)_+ - q_+ = q_1 \times q_2$$

may result in different values of q_2, depending on the choice of q_1.

Defining the product $\{q_0, q_1\}$ to be $\int_\Omega q_0(\mathbf{x}) \times q_1(\mathbf{x})\, d\mathbf{x}$ we arrive at conclusions similar to Lemmas 1 and 2.

LEMMA 3. *The Fréchet derivative of* $Q(q) = \int_\Omega (q \times \mathrm{e})\, d\mathbf{x}$ *is defined and is equal to* $\mathrm{e}(\mathbf{x})$.

The proof is elementary (See Appendix 2 for an outline).

LEMMA 4 (The Lax–Milgram theorem). *If* $Q(q_0, q_1)$ *is a bilinear quaternion form linearly depending on* $q_0(\mathbf{x})$, *then* Q *is of the form* $\{Lq_0, q_1\}$.

For the proof, see Appendix 4.

We are now ready to recognize some well-known equations of physics in the framework of differentiation of quaternion products.

2. QUATERNION REPRESENTATION OF SOME OPERATOR EQUATIONS OF PHYSICS IN THE FORM $T = A^*A$

a. *The 3-Dimensional Laplace Operator*

We introduce the operator

$$A = \mathbf{1} \cdot 0 + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K}\frac{\partial}{\partial z}$$

and its formal adjoint

$$A^* = \mathbf{1} \cdot 0 - \mathbf{i}\frac{\partial}{\partial x} - \mathbf{j}\frac{\partial}{\partial y} - \mathbf{K}\frac{\partial}{\partial z}.$$

Then AA^* is negative definite and is given by

$$AA^* = \Delta_3 = \mathbf{1}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right).$$

If we feel that AA^* should be positive definite, rather than negative definite, we define accordingly

$$A = \mathbf{1} \cdot 0 + \mathbf{i}\left(\frac{i\partial}{\partial x}\right) + \mathbf{j}\left(\frac{i\partial}{\partial y}\right) + \mathbf{K}\left(\frac{i\partial}{\partial z}\right).$$

Then $A^* = A$ and $AA^* = -\Delta_3$. Here, as usual, $i^2 = -1$. Fir purposes of complex analysis it is more convenient to rewrite the two-dimensional Laplace equation in the formalism:

$$h = \mathbf{1}\frac{\partial}{\partial t} + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K} \cdot 0,$$

$$\bar{h} = -\mathbf{1}\frac{\partial}{\partial t} - \mathbf{i}\frac{\partial}{\partial x} - \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K} \cdot 0,$$

where the operator

$$h\bar{h} = \bar{h}h = \mathbf{1}\left(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

is applied to functions of the form $u(x, y) + iv(x, y)$, the first term being utterly superfluous. On the other hand, applying h to a quaternion of the form

$$q = \mathbf{1}c_0 + \mathbf{i}f_1(x, y) + \mathbf{j}f_2(x, y) + \mathbf{K} \cdot 0,$$

we obtain

$$hq = \mathbf{1}\left(\frac{\partial}{\partial x}f_1(x, y) + \frac{\partial}{\partial y}f_2(x, y)\right) + \mathbf{K}\left(\frac{\partial}{\partial x}f_2(x, y) - \frac{\partial}{\partial y}f_1(x, y)\right).$$

If $\overline{f_1(x, y) + if_2(x, y)}$ is analytic, we have $hq \equiv 0$, i.e., h acts like the $\bar{\partial}$ operator.

b. *The Beam Equation*

We define the operator

$$A = \mathbf{1}\left(i\frac{\partial}{\partial t}\right) + \mathbf{i}\left((D(x, y))^{1/2}\frac{\partial^2}{\partial x^2}\right), \qquad D(x, y) > 0;$$

then

$$A^* = \mathbf{1}\left(i\frac{\partial}{\partial t}\right) + \mathbf{i}\left(\frac{\partial^2}{\partial x^2}D(x, y)^{1/2}\right) \tag{b1}$$

and

$$A^*A = \mathbf{1}\left[\left(-\frac{\partial^2}{\partial t^2}\right) - \frac{\partial^2}{\partial x^2}\left(D(x, y)\frac{\partial^2}{\partial x^2}\right)\right], \tag{b2}$$

which is the classical Lagrange operator. It is an easy exercise to check that if the boundary conditions are natural then A^* is the true adjoint of A, and $A^*A = AA^*$, even though the operator AA^* looks strange at a first glance.

Note. We have omitted the $\mathbf{j}$, $\mathbf{K}$ terms, but clearly we could write

$$A = \mathbf{1}(i\partial/\partial t) + \mathbf{i}((D(x, y)^{1/2})\partial^2/\partial x^2) + \mathbf{j} \cdot 0 + \mathbf{K} \cdot 0.$$

c. *The Klein–Gordon Operator*

We introduce the operator

$$A = \mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K}\frac{\partial}{\partial z} \tag{c1a}$$

and its adjoint

$$A^* = \mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) - \mathbf{i}\frac{\partial}{\partial x} - \mathbf{j}\frac{\partial}{\partial y} - \mathbf{K}\frac{\partial}{\partial z}; \tag{c1b}$$

clearly

$$AA^* = A^*A = -\partial^2/\partial(ct)^2 + \Delta = \Box. \tag{c2}$$

This decomposition also leads to a novel representation of Maxwell's equations.

3. Some General Remarks on Applications to Quantum Mechanics

It is interesting to observe that the quaternion notation has several advantages. From the Hamiltonian decomposition of $AA^*\mathbf{u} = \mathbf{f}$, we derive separate equations of a generalized system

$$A^*\psi = \phi, \qquad A\phi = \mathbf{f},$$

where, in the quantum mechanical formalism,

$$\mathbf{f} = m^2c^2\Psi.$$

A slight modification gives us

$$\begin{aligned} A &= i\hbar\left[\mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K}\frac{\partial}{\partial z}\right]. \\ A^* &= i\hbar\left[\mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) - \mathbf{i}\frac{\partial}{\partial x} - \mathbf{j}\frac{\partial}{\partial y} - \mathbf{K}\frac{\partial}{\partial z}\right], \end{aligned} \tag{c3}$$

and the canonical equations of Hamilton in the form

$$\begin{aligned} A^*\psi &= \partial W/\partial\phi, \\ A\phi &= \partial W/\partial\psi, \end{aligned} \tag{c4}$$

with

$$W = \{\phi, \phi\} + \{f, \psi\} \tag{c5}$$

defining the Hamiltonian.

The corresponding Lagrangian is given by

$$\mathscr{L} = W - \{A^*\psi, \phi\}. \tag{c6}$$

So far ϕ, ψ are undefined quaternions, which can be easily identified with the solutions of the Dirac equation, if another "factor" mc is introduced into the operators A and A^*. If mc is regarded as a scalar (i.e., $mc = \mathbf{1}(mc) + \mathbf{i}\cdot 0 + \mathbf{j}\cdot 0 + \mathbf{K}\cdot 0$), no complications arise with respect to formal adjoints of mc

($\overline{mc} \equiv mc$). At this point it natural to regard mass as a general quaternion arriving in a natural way at mass plus "spin up" or "spin down" states as conjectured by Edmonds [1], i.e., we replace m by $\mathbf{m}$.

One could form several conjectures regarding quantum mechanical interpretation of Eqs. (c4)–(c6) with mc defined in the general quaternionic form. System (c4) can be rewritten as

$$\begin{aligned} i\hbar\bar{\partial}\psi &= \mathbf{m}c \times \phi, \\ i\hbar\partial\phi &= \mathbf{m}c \times \psi, \\ \bar{\partial} \equiv \bar{A}, &\qquad \partial \equiv A, \end{aligned} \tag{c7}$$

to bring them into a more familiar appearance of Dirac equations.

We postpone this discussion, preferring to treat the Dirac equation separately (Section f). It follows easily from Vaĭnberg's theorem that systems (c4) or (c7) represent a critical point of the Lagrangian, corresponding to dual variational principles, allowing one to introduce completely different computational techniques, based on Noble's two sided variational inequalities (See [4]).

d. *Maxwell's Equations*

Let the electric current and charge density be represented by the quaternion

$$Q_e = 4\pi i[\mathbf{1}(i\rho c) + \mathbf{i}j_x + \mathbf{j}j_y + \mathbf{K}j_z] \tag{d1}$$

and the magnetic (monopole) current and magnetic charge density be represented by the quaternion

$$Q_m = 4\pi[\mathbf{1}(i\bar{\rho}c) + \mathbf{i}\bar{j}_x + \mathbf{j}\bar{j}_x + \mathbf{K}\bar{j}_z]. \tag{d2}$$

Now using the properties of the operator A, and applying it to the complex vector $\mathbf{E} + i\mathbf{H}$, we have

$$\begin{aligned} &\left[\mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{K}\frac{\partial}{\partial z}\right](\mathbf{E} + i\mathbf{H}) \\ &= \left(-\frac{\partial E_0}{\partial(ct)} - \frac{\partial E_x}{\partial x} - \frac{\partial E_y}{\partial y} - \frac{\partial E_z}{\partial z}\right)\mathbf{1} + \left(\frac{\partial E_0}{\partial x} - \frac{\partial H_x}{\partial(ct)} - \frac{\partial E_y}{\partial z} + \frac{\partial E}{\partial y}\right)\mathbf{i} \\ &\quad + \left(\frac{\partial E_0}{\partial y} + \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} - \frac{\partial H_y}{\partial(ct)}\right)\mathbf{j} + \left(\frac{\partial H_z}{\partial(ct)} + \frac{\partial E_y}{\partial x} + \frac{\partial E_x}{\partial y} + \frac{\partial E_0}{\partial z}\right)\mathbf{K} \\ &\quad + i\left\{\left(\frac{\partial H_0}{\partial(ct)} - \frac{\partial H_x}{\partial x} - \frac{\partial H_y}{\partial y} - \frac{\partial H_z}{\partial z}\right)\mathbf{1}\right. \\ &\quad + \left(\frac{\partial E_x}{\partial(ct)} + \frac{\partial H_0}{\partial x} + \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial E_y}{\partial(ct)} - \frac{\partial H_z}{\partial x} + \frac{\partial H_0}{\partial y} + \frac{\partial H_x}{\partial z}\right)\mathbf{j} \\ &\quad \left. + \left(\frac{\partial E_z}{\partial(ct)} + \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} + \frac{\partial H_0}{\partial z}\right)\mathbf{K}\right\}. \end{aligned} \tag{d3}$$

Equating $A(\mathbf{E} + i\mathbf{H})$ with $(1/c)(Q_e + iQ_m)$ we obtain the usual set of Maxwell's equations, after setting all space derivatives of E_0 and H_0 equal to zero.

However, if we assume that the quadruples (ρ, j_x, j_y, j_z), $(\bar{\rho}, \bar{j}_x, \bar{j}_y, \bar{j}_z)$ can be derived from a potential, the following argument gives another version of Maxwell's equations. We introduce the usual electric and magnetic vectors

$$\mathbf{E} = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} H_x \\ H_y \\ H_z \end{pmatrix}$$

as well as a fourth component E_0, H_0, such that

$$\begin{aligned} \frac{\partial E_0}{\partial(ct)} &= -4\pi\rho, \\ \frac{\partial E_0}{\partial x} &= -\frac{4\pi}{c}\dot{j}_x, \\ \frac{\partial E_0}{\partial y} &= -\frac{4\pi}{c}\dot{j}_y, \\ \frac{\partial E_0}{\partial z} &= -\frac{4\pi}{c}\dot{j}_z, \end{aligned} \tag{d4}$$

where, as before, ρ is the charge density, while

$$J = \begin{pmatrix} \dot{j}_x \\ \dot{j}_y \\ \dot{j}_z \end{pmatrix}$$

is the electric current. Similarly, we conjecture the possibility that there exists an H_0 such that

$$\begin{aligned} \frac{\partial H_0}{\partial(ct)} &= -4\pi\bar{\rho}, \\ \frac{\partial H_0}{\partial x} &= -\frac{4\pi}{c}\bar{j}_x, \\ \frac{\partial H_0}{\partial y} &= -\frac{4\pi}{c}\bar{j}_y, \\ \frac{\partial H_0}{\partial z} &= -\frac{4\pi}{c}\bar{j}_z, \end{aligned} \tag{d5}$$

then $A(\mathbf{E} + i\mathbf{H}) = 0$ is the set of Maxwell's equations.

It is straightforward computation that

$$A^*A(\mathbf{E} + i\mathbf{H}) \equiv 0$$

or

$$\square(\mathbf{E} + i\mathbf{H}) \equiv 0,$$

whether the assumptions concerning the existence of E_0, H_0 satisfying (d4), (d5) are true or false.

e. *Rigid Body Mechanics*

If we adopt the Euler angles θ, ϕ, ψ as the generalized coordinates describing a motion of a rigid body with a fixed point, the equations of motion are derived from Hamilton's principle of least action

$$\delta \int_{t_0}^{t_1} (T - V)^{1/2}\, ds = 0. \tag{e1}$$

The kinematic metric g^{ij} is defined by the quadratic from of the kinetic energy

$$T = \tfrac{1}{2}(g^{ij}\dot{q}_i\dot{q}_j). \tag{e2}$$

For a spinning top this means, unfortunately, that the metric g^{ij} is of the form

$$g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \cos\theta \\ 0 & \cos\theta & 1 \end{pmatrix}, \tag{e3}$$

i.e., the path described by a geodesic (e1) has a locally nonorthogonal coordinate system. This is hardly surprising, since the local map $(\theta, \phi, \psi) \to (x, y, z)$,

$$\begin{aligned} z &= \cos\theta, \\ y &= -\sin\theta\cos\phi, \\ x &= \sin\theta\sin\phi, \end{aligned} \tag{e4}$$

satisfies the constraint

$$x^2 + y^2 + z^2 = 1, \tag{e5}$$

i.e. the motion is restricted to the surface of a unit sphere, and we can not produce a global orthogonal system restricted to the surface of a unit sphere in $\mathbf{R}^3$.

Alternate sets of variables have been proposed by various authors, for example, Sansò [5] or Lattman, the general idea being to embed the problem in a four-dimensional space, so that the motion of a rigid body takes place in a three-dimensional subspace equipped with a global orthogonal system of coordinates.

Following Sansò we introduce coordinates q_0, q_1, q_2, q_3 (which may be complex valued) such that

$$\sum_{i=0}^{3} q_i^2 = 1. \tag{e6}$$

We simply postulate that the kinematic metric is Euclidean (and now we get away with it), i.e.,

$$ds^2 = C^2\left(\sum_{i=0}^{3} (dq_i)^2\right), \tag{e7}$$

where C is a real number which may be chosen later for the sake of physical convenience. Sansò and Evans have discovered that writing the equations of the rotation vector in terms of the four quantities q_i, we have the following representation:

$$\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \\ 0 \end{pmatrix} = A \begin{pmatrix} \dot{q}_3 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_0 \end{pmatrix} \tag{e8}$$

with

$$A = 2\begin{bmatrix} -q_3 & q_0 & q_2 & q_1 \\ -q_0 & -q_3 & -q_1 & q_2 \\ -q_1 & -q_2 & q_0 & -q_3 \\ q_2 & -q_1 & q_3 & q_0 \end{bmatrix}. \tag{e9}$$

The constraint (e6) becomes:

$$\sum_{i=0}^{3} q_i\dot{q}_i = 0. \tag{e10}$$

The quantities q_i can be regarded as components of the quaternion $Q = \mathbf{1}q_0 - \mathbf{i}q_1 - \mathbf{j}q_2 - \mathbf{K}q_3$.

We notice the following facts. The matrix A is orthogonal. However, in the corresponding three-dimensional representation, the rotation matrix assumes the following form:

$$B = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 + q_3^2 & -2(q_0q_3 - q_2q_1) & 2(q_0q_2 - q_3q_1) \\ 2(q_0q_3 + q_2q_1) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & -2(q_0q_1 - q_2q_3) \\ -2(q_0q_2 - q_1q_3) & 2(q_0q_1 + q_2q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix},$$

which is in general nonorthogonal.

The advantages of quaternionic representation become apparent.

The required variational representation is an easy consequence of the orthogonality of A. Since A^*A is positive if A is of the form $a_0 + i(a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{K})$

where the a_i are real, we can adopt a convention that rotations will be represented by quaternions of that form. This does not violate any previously made assumptions, and is exactly the representation of Sansò [5]. Denoting $(A^*A)^{1/2}$ by $\mathscr{A}$, we can rewrite the equations of motion for the rigid body in the form

$$\mathscr{A}y = \omega,$$

where

$$\mathscr{A} = T^*T,$$

or, factoring $\mathscr{A}$ out,

$$Ty = p,$$
$$T^*p = w,$$

which is a standard canonical form (See Noble [4] or Arthurs [7]).

In classical mechanics p does not have an obvious physical interpretation. We could refer to it as the $\frac{1}{2}$ spin, or simply as the generalized momentum conjugate to the generalized coordinate $y(t)$, following the Legendre transformation

$$p = \partial\mathscr{L}/\partial(Ty).$$

Comment. Evans and Sansò defined their quaternionic coordinates in terms of the Euler angles (within a sign change) as follows:

$$\begin{aligned} q_0 &= \cos(\theta/2)\cos(\phi + \psi)/2, \\ q_1 &= \sin(\theta/2)\cos(\phi - \psi)/2, \\ q_2 &= \sin(\theta/2)\cos(\phi - \psi)/2, \\ q_3 &= \cos(\theta/2)\sin(\phi + \psi)/2. \end{aligned}$$

f. *Dirac Equation, van der Waerden Equations*

Let α be an operator represented by components

$$\alpha_x = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\alpha_y = \begin{bmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{bmatrix}, \qquad \text{(f1)}$$

$$\alpha_z = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix};$$

let β be the matrix

$$\beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{f2}$$

Then the equation

$$i\hbar\partial\psi/\partial t = (c\boldsymbol{\alpha} \cdot \boldsymbol{\beta} + mc^2\beta)\,\psi, \tag{f3}$$

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

is the original form of Dirac's equation. It was at first regarded as a relativistic version of the Schrödinger equation.

The original form of Dirac's equation (f3) can be manipulated into various alternative forms. For example, defining matrices

$$\gamma^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \gamma = \begin{bmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{bmatrix}, \tag{f4}$$

where the σ_i are the Pauli matrices ($i = 1, 2, 3$), the Dirac equation may be written as a pair of equations

$$\begin{aligned} \hbar(i/m)\,(i\partial/\partial(ct) - \nabla \cdot \sigma)\,\chi^L &= \chi^R, \\ \hbar(i/m)\,(i\partial/\partial(ct) + \nabla \cdot \sigma)\,\chi^R &= \chi^L. \end{aligned} \tag{f5}$$

By eliminating either χ^L or χ^R, one obtains the van der Waerden equations

$$\begin{aligned} -(1/m^2)\,\Box\,\chi^L(x) &= \chi^L, \\ -(1/m^2)\,\Box\,\chi^R(x) &= \chi^R. \end{aligned} \tag{f6}$$

We define the operators

$$\Delta = \frac{i\hbar}{m}\left[\mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) - \mathbf{i}\left(\frac{\partial}{\partial x}\cdot\sigma_1\right) - \mathbf{j}\left(\frac{\partial}{\partial y}\cdot\sigma_2\right) - \mathbf{K}\left(\frac{\partial}{\partial z}\cdot\sigma_3\right)\right],$$

$$\Delta^* = \frac{i\hbar}{m}\left[\mathbf{1}\left(i\frac{\partial}{\partial(ct)}\right) + \mathbf{i}\left(\frac{\partial}{\partial x}\cdot\sigma_1\right) + \mathbf{j}\left(\frac{\partial}{\partial y}\cdot\sigma_2\right) + \mathbf{K}\left(\frac{\partial}{\partial z}\cdot\sigma_3\right)\right].$$

This enables us to rewrite the Dirac equations (f5) in the generalized canonical form

$$A\chi^L = \chi^R = \partial H/\partial \chi^R,$$
$$A^*\chi^R = \chi^L = \partial H/\partial \chi^L \tag{f7}$$

with the Hamiltonian given by

$$H = \tfrac{1}{2}[\{A\chi^L, \chi^R\} + \{\chi^L, A^*\chi^R\}]. \tag{f8}$$

4. A Theory of Critical Points of Quaternionic Functionals

The theory derived here parallels the known results summarized in Vaînberg's monograph [6]. The difficulties which arise are of a purely algebraic nature. The arguments of purely analytic nature can be reproduced without many changes from the well known texts. For an excellent review of some topics in applications of Fréchet differentiation we recommend the article by Nashed [9], the original reports of Noble [4], and the monograph by Arthurs [7].

Notation. $\mathbf{q}(x) \in L_2^{(4)}(\Omega)$ denotes that each component of $\mathbf{q}(x) = c_0(x)\mathbf{1} + c_1(x)\mathbf{i} + c_2(x)\mathbf{j} + c_3(x)\mathbf{K}$ is an $L_2(\Omega)$ function. The norm of $\mathbf{q}(x)$ is $\|\mathbf{q}(x)\| = \{\sum_{i=0}^{b} \| c_i(x)\|^2\}^{1/2}$. With this norm topology we define continuity of a map from $L_2^{(4)}(\Omega)$ to the scalar quaternion field by requiring that each component of $\mathbf{Q}(\mathbf{q}(x))$ is continuous, that is, $\mathbf{Q}(q(x)) = a_0\mathbf{1} + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{K}$ is a continuous functional of $\mathbf{q}(x)$ if each element of $\mathbf{Q}(\mathbf{q}(x))$ continuously depends on $\mathbf{q}(x)$. This implies that whenever $\{q_i(x)\} \to^{L_2^{(4)}(\Omega)} \bar{\mathbf{q}}(x)$ then the corresponding sequences of numbers converge: $a_i(q_n(x)) \to^{L_2(\Omega)} a_i(\bar{\mathbf{q}}(x))$, $i = 0, 1, 2, 3$. We shall consider quite arbitrary functionals generally mapping quaternions $q_1(x), q_2(x), \ldots, q_n(x)$ into a constant quaternion $\mathbf{Q}$.

The simplest case involves a bilinear, or a sequilinear form

$$\mathbf{Q} = \{q_1, q_2\} \overset{\text{def}}{=} \int_\Omega (q_1(x) \times \overline{q_2(x)})\, dx,$$

i.e.,

$$\mathbf{Q} = \tfrac{1}{2}Q_0 + \mathbf{i}Q_1 + \mathbf{j}Q_2 + \mathbf{K}Q_3$$
$$= \tfrac{1}{2}\int_\Omega a_0(x)\, dx + \mathbf{i}\int_\Omega a_1(x)\, dx + \mathbf{j}\int_\Omega a_2(x)\, dx + \mathbf{K}\int_\Omega a_2(x)\, dx,$$

where, for almost each $x \in \Omega$, $\mathbf{1}a_0(x) + \mathbf{i}a_1(x) + \mathbf{j}a_2(x) + \mathbf{K}a_3(x) = q_1(x) \times q_2(x)$. In general the product $\{\ ,\ \}_\Omega$ will be defined by the formula $\{q_1, q_2\} = \int_\Omega (q_1(x) \times \overline{q_2(x)})\, dx$—where as before $\times$ denotes quaternionic multiplication and the bar denotes conjugation. The sequilinear property of $\{q_1, q_2\}_\Omega$ arises naturally. We postulate $c\mathbf{Q}(q_1, q_2) = \mathbf{Q}(cq_1, q_2) = \mathbf{Q}(q_1, \bar{c}q_2)$ to ensure the usual inner product property of $\{q_1, q_2\}$ if the **i**, **j**, and **K** components of q_1 and q_2 are equal to zero.

Unfortunately we have to distinguish between right and left Fréchet derivatives. The rules of differentiation are introduced as follows.

The identities

$$\mathbf{Q}(q_1 - \epsilon\eta) - \mathbf{Q}(q_1) = \epsilon\{\eta, \phi\} + o(\epsilon^2)$$

$$\mathbf{Q}(q_1 + \epsilon\eta) - \mathbf{Q}(q_1) = \epsilon\{\overline{\psi, \eta}\} + o(\epsilon^2)$$

define, respectively, the left and right Fréchet derivatives ϕ, ψ, whenever they are satisfied.

If **Q** has only one nonzero component (assume without any loss of generality that **i**, **j**, **K** components are equal to zero), then the sequilinear property of the product $\{\ ,\ \}$ implies $\psi = \phi$. Hence Fréchet differentiation of sequilinear complex- or real-valued quaternionic functionals is uniquely defined. So far in mathematical physics we have only encountered the case when the functionals were either real or complex valued. However, the general theory requires considerable discussion, carefully separating the right and left differentiation.

For this reason we shall denote the right and left factors of sequilinear quaternionic products by $|\, q_2\}$ and $\{q_1\,|$, respectively, in a manner borrowed from Dirac's bra-Ket notation. Then

$$\frac{\partial \mathbf{Q}}{\partial\{q_1|} = |q_2\}, \qquad \frac{\partial \mathbf{Q}}{\partial|q_2\}} = \{q_1|$$

implies the correct left or right multiplication in the reconstruction of the original functional **Q**. Moreover we can borrow the physics notation and denote

$$\{Aq_1, q_2\} \quad \text{by} \quad \{q_1 \mid A \mid q_2\}.$$

This is equal by the definition of A^* to $\{q_1, A^*q_2\}$. The "right" or "left" definitions are promptly ignoted if the functional is complex or real valued. For the scalar case, when the quaternion-valued functional possesses only one nonzero component, the entire Fréchet differentiation theory as outlined by Vaînberg [6] can be reproduced with only minor modifications. For the general case this theory needs to be rederived. We shall start this project with a technical lemma.

LEMMA 1. *Let $Q(q(x))$ be a functional possessing the left Fréchet derivative. Then it possesses the right Fréchet derivative and the critical points of Q with respect to left differentiation are the critical points with respect to the right differentiation. (A critical point* e *is the point at which the derivative vanishes.)*

Proof (a direct computation). In what follows we may consider quaternion valued functionals of n-copies of $L_2^{(4)}(\Omega)$, which may be considered a product space $L_{2,n}^{(4)}(\Omega) = L_2^{(4)}(\Omega) \times L_2^{(4)}(\Omega) \cdots L_2^{(4)}(\Omega)$ with an obvious definition of a quaternionic n-product $\mathbf{q}^{(1)} \times \mathbf{q}^{(2)} = \{q_1^{(1)} \times q_1^{(2)}, q_2^{(1)} \times q_2^{(2)} \cdots q_n^{(1)} \times q_n^{(2)}\}$, where $\mathbf{q}$ is the n-tuple $\mathbf{q} = \{q_1, q_2, \ldots, q_n\}$, $\mathbf{q}_1 \in L_2^{(4)}(\Omega)$, $\mathbf{q}_2 \in L_2^{(4)}(\Omega), \ldots$, etc. Left and right Fréchet differentiability is defined in an analogous manner.

Before proving the basic Theorem 1, we need to introduce some definitions.

DEFINITION. A curve in $L_{2,n}^{(4)}(\Omega)$ is a continuous mapping from $\mathbf{R}$ into $L_2^{(4)}(\Omega)$. (Continuity is defined with respect to $L_2^{(4)}(\Omega)$ norm.) A line passing through a point $\tilde{\mathbf{q}}$ in the direction of $\eta \in L_{2,n}^{(4)}(\Omega)$ is the collection of all points of the form $q + t\eta$, $-\infty < t < +\infty$, $q = q_1, q_2, \ldots, q_n$.

LEMMA 2. *Suppose that $\Phi(\mathbf{q}(x))$ is a real-valued functional defined for all $\mathbf{q} \in L_{2,n}^{(4)}(\Omega)$. Then a necessary and sufficient condition for $\Phi(\mathbf{q}(x))$ to attain a local minimum (maximum) at $\mathbf{q} = \tilde{\mathbf{q}}$ is that $\Phi(\mathbf{q})$ attains a local minimum (maximum) at $\tilde{\mathbf{q}}$ on any line passing through $\mathbf{q}$ in $L_{2,n}^{(4)}(\Omega)$.*

Proof. This follows almost directly from the Bing–Anderson theorem on the structure of a separable Hilbert space.

COROLLARY. *$\Phi(\mathbf{q}(x))$ attains a local minimum at $\mathbf{q} = \tilde{\mathbf{q}}$ if it attains a local minimum on any curve passing through $\tilde{\mathbf{q}}$.*

THEOREM 1. *Let us consider a real-valued functional $\Phi(q)$, $q \in L_{2,n}^{(4)}(\Omega)$, possessing Fréchet derivatives in some neighborhood of $\tilde{\mathbf{q}} \in L_{2,n}^{(4)}(\Omega)$. A necessary condition for a local extremum of $\Phi(\mathbf{q})$ at $\tilde{\mathbf{q}}$ is the vanishing at $\tilde{\mathbf{q}}$ of the left (or right) gradient of Φ.*

Proof (almost trivial). Suppose that $\mathbf{0}$ is a local extremum of Φ; then $\Phi(\tilde{\mathbf{q}} + \ + t\eta) - \Phi(\mathbf{q}) = t\{\eta, \partial\Phi/\partial\mathbf{q}\} + o(t^2)$ and for sufficiently small $|t|$ if $\partial\Phi/\partial\mathbf{q} \neq \varnothing$, we can choose η such that $\{\eta, \partial\Phi/\partial\mathbf{q}\}_{(n)}$ is positive (or is negative) denying the fact that $\tilde{\mathbf{q}}$ was an extremal point of Φ.

Note 1. Clearly this is not a sufficient condition for an extremum of Φ.

Note 2. The right or left gradient of Φ is not generally real- or complex-valued in any region of $L_{2,n}^{(4)}(\Omega)$ even if Φ is.

AN EXAMPLE. The Hamiltonian (f7) generates the corresponding Lagrangian

$$\mathscr{L} = H - \{A\chi^L, \chi^R\}.$$

If $\{A\chi^L, \chi^R\}$ is a scalar we need not distinguish between right and left derivatives, and the vanishing of the derivative of the Lagrangian is equivalent to the existence of a solution of the system of equations (f7). If $\{A\chi^L, \chi^R\}$ has more than one nonvanishing quaternionic component, clearly this simple analysis is no longer applicable.

5. A Class of Boundary Value Problems

Let T be a quaternion operator of the form

$$T = \tau_0 \mathbf{1} + \tau_1 \mathbf{i} + \tau_2 \mathbf{j} + \tau_3 \mathbf{K},$$

where τ_0, τ_1, τ_2, τ_3 are operators mapping a Hilbert space H_ϕ into a Hilbert space H_ϕ. (Note: H_ϕ may be actually a direct product of Hilbert spaces: $H_\phi = \sum_{i=1}^{4} H_{\phi_i}$, with τ_i acting only on H_{ϕ_i}, $i = 1, 2, 3, 4$, but this is an unimportant technical detail, which will be ignored here.) T^* denotes the quaternionic adjoint of T in the same sense as in Sections 3 and 4, i.e., for any quaternions $\phi \in H_\phi{}^4$, $\psi \in H_\psi{}^4$ ($H^4 = H \oplus H \oplus H \oplus H$), the following equality holds

$$\{T\phi, \psi\}_{H_\psi{}^4} = \{\phi, T^*\psi\}_{H_\phi{}^4},$$

for all $\phi \in H_\phi{}^4$, $\psi \in H_\psi{}^4$.

Suppose that Hilbert spaces H_{ϕ_i}, H_{ψ_i} are spaces of functions confined to regions Ω_1, Ω_2 of $\mathbf{R}^n$, $\mathbf{R}^m$, respectively. For the sake of simplicity we shall take $n = m$ and $\Omega_1 = \Omega_2 = \Omega$, with $H_{\phi_i} = H_{\psi_i} = \phi_{\phi_j} = H_{\psi_j}$, $i, j = 1, 2, 3, 4$. The boundary of Ω will be denoted by $\partial\Omega$, and $\Omega \cup \partial\Omega$ by $\bar{\Omega}$.

Again T and T^* are quaternionic operators such that the products $\{T\phi, \psi)_{H_\psi{}^4}$, $\{\phi, T^*\psi\}_{H_\phi{}^4}$ are real valued and are related by the following equality

$$\{T\phi, \phi\}_{H_\psi{}^4(\Omega)} = \{\phi, T^*\psi\}_{H_\phi{}^4(\Omega)} + B(\phi, \psi)_{\partial\Omega}, \tag{4.1}$$

where $B(\phi, \psi)_{\partial\Omega}$ is a number which only depends on the behavior of the components of ϕ and ψ on $\partial\Omega$.

Moreover we shall postulate that there exists a linear map σ such that all components of quaternions $\sigma\phi$, ψ restricted to $\partial\Omega$ are square integrable, and the product

$$\{\sigma\phi, \psi\}_{\partial\Omega} = \int_{\partial\Omega} (\sigma\phi \times \bar{\psi})\, ds$$

exists and is real valued for all $\psi \in H_\phi{}^4$, all $\psi \in H_\psi{}^4$. Then we can define the adjoint operator σ^* by the equality

$$\int_{\partial\Omega} (\sigma\phi \times \bar{\psi})\, dx = \int_{\partial\Omega} (\psi \times \overline{\sigma^*\psi})\, dx.$$

We shall consider a class of boundary value problems.

$$T\phi = \psi \in H_\psi{}^4(\Omega), \tag{4.2a}$$

$$T^*\psi = g \in H_\phi{}^4(\Omega), \tag{4.2b}$$

$$\begin{aligned} \sigma\phi\,|_{\partial\Omega} &= \mu(x)|_{x\in\partial\Omega}\,, \\ \phi^*\phi\,|_{\partial\Omega} &= 0. \end{aligned} \tag{4.3}$$

Assuming that all products are real valued, the existing critical point theory can be applied immediately to this class of problems. See [6, 7, 9]. The corresponding Hamiltonian and Langrangian are given, respectively, by

$$H = \tfrac{1}{2}\{\psi, \psi\}_{H_\psi{}^4(\Omega)} + \{\phi, g\}_{H_\phi{}^4(\Omega)}\,. \tag{4.4}$$

The Lagrangian is given by

$$\begin{aligned} &\{T\phi, \psi\}_{H_\psi{}^4(\Omega)} - \tfrac{1}{2}\{\psi, \psi\}_{H_\phi{}^4(\Omega)} - \{\phi, g\} + \{(\sigma\phi - \mu), \psi\}_{\partial\Omega} \\ &\quad = \{T\phi, \psi\}_{H_\psi{}^4(\Omega)} - H + \{(\sigma\psi - \mu), \psi\}_{\partial\Omega} \\ &\quad = \{T\phi, \psi\}_{H_\psi{}^4(\Omega)} - H - \bar{H}. \end{aligned}$$

These equations can be written in the following form:

$$\begin{aligned} \{T\phi, \psi\}_{H_\psi{}^4} &= \left\{\frac{\partial H}{\partial\psi}, \psi\right\}_{H_\psi{}^4} \quad \text{in} \quad \Omega \\ \{\psi, T^*\psi\}_{H_\phi{}^4} &= \left\{\frac{\partial H}{\partial\phi}, \phi\right\}_{H_\phi{}^4} \quad \text{in} \quad \Omega \end{aligned} \tag{4.5}$$

$$\begin{aligned} \{(\sigma\phi - \mu), \psi\}_{\partial\Omega} &= \left\{\frac{\partial H}{\partial\psi}, \psi\right\}_{\partial\Omega} \quad \text{in} \quad \partial\Omega \\ \{\phi, \sigma^*\psi\}_{\partial\Omega} &= \left\{\phi, \frac{\partial H}{\partial\phi}\right\}_{\partial\Omega} \quad \text{in} \quad \partial\Omega. \end{aligned} \tag{4.6}$$

The following variational problems can be formulated.

If L is convex in ϕ and concave in ψ in a neighborhood of a point $z_0 = (\phi_0, \psi_0)$ such that

$$\frac{\partial L}{\partial\phi\,|_{z_0}} = 0, \qquad \frac{\partial L}{\partial\psi\,|_{z_0}} = 0,$$

then the point z_0 is a min-max point of L. We need to comment that aside from conditions $\phi \in H^4$, $\psi \in H^4$, the functions ϕ, ψ need to be continuous in some neighborhood of the boundary $\partial\Omega$, in some sense, otherwise the boundary value problems lose all physical meaning.

Concluding Remarks

In the case of real-valued functionals major parts of known boundary value theories may be routinely rederived with changes effected in only the algebraic rules. See, for example, [10] or [11]. Moreover, as was shown by Gurtin [12] the inner products for quaternionic components in the formulation of the $L_2^{(4)}(\Omega)$ theory can be replaced by convolution products or other bilinear forms whenever the operator occurring in the physical problem acquires symmetry with respect to the new bilinear form.

The numerical bounds and estimates similar to [11] are deliberately omitted at this stage of research. However, arguments parallel to Arthurs [7, Chap. 3] with computation of upper and lower bounds of Lagrangian functionals should be derived for specific quaternionic formulations of physical problems to establish the usefulness of the general theory.

APPENDIX 1: The Quaternion Algebra

The quaternion units $\{\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{K}\}$ obey the multiplication table:

	1	**i**	**j**	**K**
1	**1**	**i**	**j**	**K**
i	**i**	$-\mathbf{1}$	**K**	$-\mathbf{j}$
j	**j**	$-\mathbf{K}$	$-\mathbf{1}$	**i**
K	**K**	**j**	$-\mathbf{i}$	$-\mathbf{1}$

The usual (componentwise) addition rules are postulated for quaternions. Ring structure is assigned with respect to the operations of addition and multiplication.

Some algebraic properties of quaternions are listed below.

(a) Multiplication is associative; i.e.,

$$\mathbf{q}_1 \times (\mathbf{q}_2 \times \mathbf{q}_3) = (\mathbf{q}_1 \times \mathbf{q}_2) \times \mathbf{q}_3$$

(obviously it is not commutative).

(b) Division is defined, except division by the zero quaternion

$$\varnothing = \mathbf{1} \cdot \mathbf{0} + \mathbf{i} \cdot \mathbf{0} + \mathbf{j} \cdot \mathbf{0} + \mathbf{K} \cdot \mathbf{0}.$$

In fact, the division formula is easily derived. Given

$$\mathbf{A} = \mathbf{1}a_0 + \mathbf{i}a_1 + \mathbf{j}a_2 + \mathbf{K}a_3 \qquad (a_i \in \mathbf{R})$$

and

$$\varnothing \neq \mathbf{C} = \mathbf{1}c_0 + \mathbf{i}c_1 + \mathbf{j}c_2 + \mathbf{K}c_3; \qquad (c_i \in \mathbf{R})$$

there exist quaternions $\boldsymbol{\beta}_l$ and $\boldsymbol{\beta}_r$ such that

$$\mathbf{A} = \boldsymbol{\beta}_l \mathbf{C} = \mathbf{C} \boldsymbol{\beta}_r .$$

The proof involves the solution of four equations in four unknowns with the determinant of the coefficients being equal to

$$D = \sum_{i=0}^{3} c_i^4 + \sum_{i=0}^{3} c_i^2 \left(\sum_{j \neq i} c_j^2 \right),$$

which is nonvanishing, unless $c_1 = c_2 = c_3 = c_4 = 0$, i.e., unless $\mathbf{C} = \varnothing$ (the zero quaternion).

The vectors of the coefficients of $\boldsymbol{\beta}_l$ and $\boldsymbol{\beta}_r$ are given by

$$\boldsymbol{\beta}_l = \gamma^{-1} \alpha,$$
$$\boldsymbol{\beta}_r = (\gamma^T)^{-1} \alpha,$$

respectively, where

$$\alpha = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

$$\gamma = \begin{bmatrix} c_0 & -c_1 & -c_2 & -c_3 \\ c_1 & c_0 & c_3 & -c_2 \\ c_2 & -c_3 & c_0 & c_1 \\ c_3 & c_2 & -c_1 & c_0 \end{bmatrix},$$

and γ^T is the transpose of γ.

APPENDIX 2: FRÉCHET DERIVATIVE

We illustrate the definition of a left (right) Fréchet derivative by offering easy examples.

(a) Compute the derivative of $f(q)$ in the direction of q_2, where

$$q = a_0(x)\mathbf{1} + a_1(x)\mathbf{i} + a_2(x)\mathbf{j} + a_3(x)\mathbf{K}, \qquad x \in \mathbf{R}^n,$$

and

$$f(q) = c_1(q \times q) + c_2 q,$$

and q_2 is an arbitrary nonzero quaternion.

We compute

$$f(q + \epsilon q_2) - f(q) = \epsilon(2a_0 c_1 + c_2) q_2 .$$

Hence, the Fréchet left derivative exists and, as asserted in Lemma 2, it is a scalar $2a_0c_1 + c_2$.

(b) It takes an elementary computation to show that, for a product

$$Q = \{q_1, q_2\} = \int_\Omega (q_1 \times q_2)\, dx,$$

the Fréchet left derivative with respect to q_1 is defined and is equal to q_2. Observe that

$$Q(q_1 + \epsilon\bar{q}) - Q(q_1) = \{\epsilon\bar{q}, q_2\}.$$

EXAMPLE 1. Consider the case of the beam equation

$$Lw = AA^*w = s(x, t),$$

where $s(x, t)$ is the applied load.

The Legendre transformation takes the form:

$$A^*W = \left[\mathbf{i}\frac{\partial}{\partial\tau} + \mathbf{j}\left(\frac{\partial^2}{\partial x^2}(D(x, y))^{1/2}\frac{\partial^2}{\partial x^2}\right)\right] W = p,$$

$$Ap = \left[\mathbf{i}\frac{\partial}{\partial t} + \mathbf{j}\frac{\partial^2}{\partial x^2}(D(x, y))^{1/2}\frac{\partial^2}{\partial x^2}\right](\mathbf{i}p_0 + \mathbf{j}p_1)$$

$$= \mathbf{1}(s(x, \tau)),$$

$$A^*w = \partial H/\partial p, \qquad Ap = \partial H/\partial w,$$

where

$$H = \tfrac{1}{2}\{w, AA^*w\} = \tfrac{1}{2}\{A^*w, A^*w\}$$

$$\underset{\text{def}}{\equiv} \tfrac{1}{2}\int_0^T\int_0^l (w(x, t)\cdot s(x, t))\, dx\, dt = \tfrac{1}{2}\{p, p\}.$$

As before, $\{\ ,\ \}$ denotes the quaternionic product, while $\cdot$ is the ordinary pointwise multiplication of scalars. The apparent negative sign of the energy product $\{\ ,\ \}$ is retained in the pointwise product

$$\tfrac{1}{2}\int_0^T\int_0^l w(x, t)\cdot s(x, t)\, dx\, dt$$

since $s(x, t)$ is the negative of the force applied to the beam.

The Lagrangian $\mathscr{L}$ is given by

$$\mathscr{L} = \{A^*w, p\} - \mathscr{H} = \{w, Ap\} - \mathscr{H}.$$

A solution of these "canonical" equations coincides with the critical point $\hat{p}$, $\hat{w}$, exactly as in the classical theory.

APPENDIX 3: BILINEAR MAPS

Properties of bilinear maps

$$\int_\Omega q_1(\mathbf{x}) + q_2(\mathbf{x})\, dx = \mathbf{J}(q_1(x), q_2(x))$$

(where **J** belongs to the ring of constant quaternions).

BASIC LEMMA. *A necessary and sufficient condition for the stationary behavior of* **J** *is* $q_1 = q_2 = \varnothing$. *There are no other critical points.*

Proof. Regard $\hat{\psi}$ as fixed. Suppose that $\mathbf{J}(\hat{\phi})$ is a critical point of $\mathbf{J}(\phi, \hat{\psi})$. We perform Fréchet differentiation, i.e., we vary ϕ by substituting $\phi = \hat{\phi} + \epsilon\xi$, $\epsilon \in \mathbf{R}$, ϵ controlling the magnitude of ξ,

$$\mathbf{J}(\hat{\phi} + \epsilon\xi, \hat{\psi}) = \int_\Omega (\hat{\phi} + \epsilon\xi) \times \hat{\psi})\, dx,$$

$$\begin{aligned} \Delta\mathbf{J} = \mathbf{J}(\hat{\phi} + \epsilon\xi) - \mathbf{J}(\hat{\phi}) &= \epsilon \int_\Omega (\xi \times \hat{\psi})\, dx \\ &= \epsilon \left[\mathbf{1} \left(\int_\Omega (\xi_0 \hat{\psi}_0)\, dx \sum_{i=1}^{3} \int \xi_i \hat{\psi}_i\, dx \right) \right. \\ &\qquad \left. + \mathbf{i} \int_\Omega [(\xi_2 \hat{\psi}_3) - (\xi_3 \hat{\psi}_2)]\, dx + \cdots \right]; \end{aligned}$$

labeling

$$\epsilon \int_\Omega (\xi_0 \hat{\psi}_0)\, dx = \Delta\mathbf{J}_0,$$

etc., we obtain

$$\epsilon \left[\int_\Omega (\xi \times \hat{\psi}) \right] dx = \mathbf{1}\Delta J_{\phi_0} + \mathbf{i}\Delta J_{\phi_1} + \mathbf{j}\Delta J_{\phi_2} + \mathbf{K}\Delta J_{\phi_3}.$$

Stationary behavior of the quaternionic integral $J(\phi, \hat{\psi})$ implies stationary behavior of each component. Hence, dividing by ϵ and letting ϵ approach zero, we obtain the required result, namely $\Delta\mathbf{J}\phi = 0$ implies that for every ξ, $\int_\Omega (\xi \times \hat{\psi})\, dx = 0$, which is possible only if $\hat{\psi} = 0$.

A similar result is obtained by fixing q_1 and varying q_2.

The converse is trivial.

COROLLARY. $Aq_1 = 0$ *is a critical point of*

$$\mathbf{J}(Aq_1, q_2) = \int_\Omega (Aq_1 \times q_2)\, dx$$

and

$$Aq_1 = \partial\mathscr{H}/\partial q_2\,, \qquad A^*q_2 = \partial\mathscr{H}/\partial q_1$$

are critical points of

$$\{Aq_1\,, q_2\} - \mathscr{H}(q_1\,, q_2),$$

where $\mathscr{H}(q_1\,, q_2)$ *is an arbitrary map of quaternion-valued functions* $q_1(x)$, $q_2(x)$ *into the complex numbers under the symmetric differentiation definition.*

Note. $\partial\mathscr{H}/\partial q$ is uniquely defined if the map $\mathscr{H}$: $(Q(x) \oplus Q(x)) \to Q$ is one dimensional, i.e., it is either real or complex valued, since then the "right" and "left" derivatives coincide.

APPENDIX 4

THE RIESZ REPRESENTATION THEOREM, LAX-MILGRAM THEOREM. *Let* $\mathbf{Q}$ $(=a_0\mathbf{1} + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{K})$ *be a quaternion-valued functional* $\mathbf{Q} = \mathbf{Q}(q(\mathbf{x}))$, $\mathbf{x} \in \Omega$, *which is linear in* $q(\mathbf{x})$ *and continuous in the* $L_2^{(4)}(\Omega)$ *topology (i.e.,* $q(\mathbf{x}) \to^{L_2^{(4)}} \varnothing$ *implies* $\mathbf{Q} \to \varnothing$*), then there exists quaternions* $\phi_1(\mathbf{x}), \phi_2(\mathbf{x})$, *such that*

$$Q(q) = \int_\Omega \phi_1(\mathbf{x}) \times q(\mathbf{x})\, d\mathbf{x} = \int_\Omega q(\mathbf{x}) \times \phi_2(\mathbf{x})\, d\mathbf{x}$$

for every $q(\mathbf{x}) \in L_2^{(4)}(\Omega)$. *Moreover,* $\phi_1\,, \phi_2$ *are unique in the* $L_2^{(4)}(\Omega)$ *sense.*

Proof. We shall prove only the existence of $\phi_1(\mathbf{x})$ since the argument concerning the existence of $\phi_2(\mathbf{x})$ is identical. Since $L_2^{(4)}(\Omega)$ is a Hilbert space (observe that algebraic properties were not used in the definition), the continuous linear map $q(\mathbf{x}) \to a_i(q(\mathbf{x}))$ $(i = 0, 1, 2, 3)$ satisfies the requirements of the Riesz representation theorem.

We shall simply construct a quaternion $\phi(x)$ with the required properties.

If $\phi(\mathbf{x}) = \mathscr{C}_0(x)\mathbf{1} + \mathscr{C}_1(\mathbf{x})\mathbf{i} + \mathscr{C}_2(\mathbf{x})\mathbf{j} + \mathscr{C}_3(\mathbf{x})\mathbf{K}$

$$\int_\Omega \phi(\mathbf{x}) \times q(\mathbf{x})\, d\mathbf{x} = (\mathscr{C}_0 c_3 - \mathscr{C}_3 c_2)]\,\mathbf{i} + [(\mathscr{C}_0 c_2 + c_0\mathscr{C}_2) + (\mathscr{C}_3 c_1 - \mathscr{C}_1 c_3)]\,\mathbf{j} + [(\mathscr{C}_0 c_3 + c_0\mathscr{C}_3) + (\mathscr{C}_1 c_2 - \mathscr{C}_2 c_1)]\,\mathbf{K} = \mathbf{Q}.$$

Obviously, $\mathbf{Q}$ linearly depends on $q(\mathbf{x})$. Let us select one of the components of $\mathbf{Q}$, say $a_2 = (\mathscr{C}_0 c_2 + c_0\mathscr{C}_2) + (\mathscr{C}_3 c_1 - \mathscr{C}_1 c_3)$. a_2 does continuously depend on $q(\mathbf{x})$ regarded as a four-component vector

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

Hence by the classical version of the Riesz representation theorem there exists a vector B such that $\mathbf{Q} = \int_\Omega \sum_{i=0}^{3} (c_i B_i)\, d\mathbf{x}$. B is identified with

$$\begin{bmatrix} \mathscr{C}_0 \\ -\mathscr{C}_1 \\ \mathscr{C}_2 \\ \mathscr{C}_3 \end{bmatrix}.$$

Repeating this argument for each component of $\mathbf{Q}$, we conclude that there exists $\phi(\mathbf{x})$ satisfying the left multiplication property. Q.E.D.

An almost identical argument works in the proof of the Lax–Milgram property, which is stated as Lemma 4 in this paper. We shall not repeat the arguments.

References

1. J. D. Edmonds, Complex energies in relativistic quantum theory, *Found. Physics* **4** No. 4 (1973), 473–479.
2. D. J. Evans, On the representation of orientation space, *Molecular Phys.* **34** No. 2 (1977), 317–325.
3. L. Euler, "Novi Comment.," Sect. 6, p. 208. St. Petersburg, Russia, 1776.
4. B. Noble, "Complementary Variational Principles for Boundary Value Problems," Technical Report No. 558, M.R.C. University of Wisconsin, Madison, 1965.
5. F. Sansó, A further account of roto-translations and the use of the method of conditioned observations.
6. M. M. Vaĭnberg, "Variational Methods for Investigation of Non-linear Operators," (transl. from Russian), Holden–Day, San Francisco, 1963.
7. A. M. Arthurs, "Complementary variational principles," Oxford Mathematical Monographs, Oxford Univ. Press (Clarendon) London, 1970.
8. J. Dieudonné, "Linear algebra and geometry," Hermann, Paris and Houghton-Mefflin, Boston, 1969.
9. M. Z. Nashed, Differentiability and related properties of nonlinear operators: some aspects of the role of differentials in nonlinear functional analysis, *in* "Nonlinear Functional Analysis and Applications" (L. B. Rall, Ed.), pp. 103–309, Academic Press, New York, 1971.
10. V. Komkov, Application of Rall's theorem to classical elastodynamics, *J. Math. Anal. Appl.* **14** (1966), 511–521.
11. A. M. Arthurs, A note on Komkov's class of boundary value problems and associated variational principles, *J. Math. Anal. Appl.* **33** (1971), 402–407.
12. M. E. Gurtin, Variational principles for linear initial value problems. *Quart. Appl. Math.* **22** (1964), 252–256.