IN PRAISE OF QUATERNIONS

JOACHIM LAMBEK

With an appendix on the algebra of biquaternions
Michael Barr

Abstract. This is a survey of some of the applications of quaternions to physics in the 20th century. In the first half century, an elegant presentation of Maxwell's equations and special relativity was achieved with the help of biquaternions, that is, quaternions with complex coefficients. However, a quaternionic derivation of Dirac's celebrated equation of the electron depended on the observation that all 4×4 real matrices can be generated by quaternions and their duals.

On examine quelques applications des quaternions à la physique du vingtième siècle. Le premier moitié du siècle avait vu une présentation élégantes des equations de Maxwell et de la relativité specialle par les quaternions avec des coefficients complexes. Cependant, une dérivation de l'équation célèbre de Dirac dépendait sur l'observation que toutes les matrices 4×4 réelles peuvent être generées par les representations regulières des quaternions.

1. Prologue.

This is an expository article attempting to acquaint algebraically inclined readers with some basic notions of modern physics, making use of Hamilton's quaternions rather than the more sophisticated spinor calculus. While quaternions play almost no rôle in mainstream physics, they afford a quick entry into the world of special relativity and allow one to formulate the Maxwell-Lorentz theory of electro-magnetism and the Dirac equation of the electron with a minimum of mathematical prerequisites. Marginally, quaternions even give us a glimpse of the Feynman diagrams appearing in the standard model.

As everyone knows, quaternions were invented (discovered?) by William Rowan Hamilton. Carl Friedrich Gauss is said to have anticipated them, but did not publish. Simon Altmann makes claims for a prior discovery by Benjamin Olinde Rodrigues, a French mathematician, banker and utopian socialist.[1)] I have glanced at the article by Rodrigues and am impressed by his technical grasp of rotations, but could not spot any explicit mention of a division algebra.

Quaternions offered an early promise for applications to physics, but met a challenge when the Michelson-Morley experiment suggested the invariance of $x_0^2 - x_1^2 - x_2^2 - x_3^2$, and not that of $x_0^2 + x_1^2 + x_2^2 + x_3^2$, the norm of a quaternion. The early attempt to overcome this problem led people to look at "biquaternions", quaternions with complex coefficients. This worked neatly for Maxwell's equation and special relativity, but not for Dirac's equation, which required that the imaginary square root of -1 be replaced by a suitable matrix which anticommutes with the basic quaternion units.

Even as the quaternion approach was improved, physics evolved into quantum electrodynamics and the standard model. Still, quaternions proved useful to some extent for classifying

2010 Mathematics Subject Classification: Primary: 16A40, Secondary: 81A06.
Key words and phrases: Classification of elementary particles; Quaternionic treatment of Dirac equation; Three temporal dimensions.
© Joachim Lambek, . Permission to copy for private use granted.

the fundamental particles of the standard model and to describe Feynman diagrams. They also led Stephen Adler to study the dynamics of particles in a quaternionic Hilbert space, a subject I will not go into here.

2. Biquaternions and the grammar of special relativity.

The algebra **H** of quaternions is finitely generated over the field of real numbers by the basic units i_0, i_1, i_2 and i_3, where

$$i_0 = 1,\ i_1^2 = i_2^2 = i_3^2 = i_1 i_2 i_3 = -1.$$

(Other equations, such as $i_1 i_2 = i_3$, easily follow.)

With any quaternion

$$a = a_0 + i_1\ a_1 + i_2 a_2 + i_3 a_3$$

one associates its *conjugate*

$$a^\dagger = a_0 - i_1 a_1 - i_2 a_2 - i_3 a_3$$

and its *norm*

$$N(a) = aa^\dagger = a^\dagger a = a_0^2 + a_1^2 + a_2^2 + a_3^2;$$

and any non zero quaternion has an inverse

$$a^{-1} = a^\dagger / N(a).$$

It is sometimes convenient to consider the *scalar* and *vector* parts of the quaternion a separately, namely a_0 and

$$\mathbf{a} = i_1 a_1 + i_2 a_2 + i_3 a_3,$$

which is indeed what Oliver Heaviside did when introducing his vector calculus. We note that

$$(ab)^\dagger = b^\dagger a^\dagger,\ N(ab) = N(a)N(b).$$

Disappointingly, the norm of a quaternion is the sum of four squares, but special relativity suggests that the quadratic form

$$a_0^2 - a_1^2 - a_2^2 - a_3^2$$

should play a more prominent rôle.

The language of quaternions had evoked strong reactions among physicists. For example, Thomson (aka Kevin), Heaviside and Minkowski were violently opposed to them, while Maxwell, Dirac and Adler were (or are) strongly in favour. Maxwell died too young to formulate his own equations in the language of quaternions. This was first done in 1912, independently by Conway and Silberstein.

Originally, they worked with *biquaternions*, that is quaternions with complex coefficients. However, in retrospect, it is clear that the only rôle played by the imaginary number i was that $i^2 = -1$ and that $ii_\alpha = i_\alpha i$ for $\alpha = 1, 2$ and 3. We will see later that the same rôle can be played by a real matrix, for example by the right representation of the basic quaternion i_1. For historical reasons, I will retain the imaginary number i in this section. (Also, this decision will allow us to postpone the puzzling question of missing dimensions until later.) For a biquaternion a, we must distinguish the quaternion conjugate $a^\dagger$ from the complex conjugate a^*.

One represents a point in space-time (better: time-space) by a *Hermitian* biquaternion

$$x = x_0 + i\mathbf{x} \ (x_0 = t)$$

for which $x^* = x^\dagger$, that is $x^{*\dagger} = x$, so that

$$N(x) = t^2 - \mathbf{x} \circ \mathbf{x} = x_0^2 - x_1^2 - x_2^2 - x_3^2,$$

as suggested by special relativity. Here $\circ$ denotes the Heaviside scalar product.

I have picked units so that the speed of light $c = 1$, Planck's constant $h = 2\pi$ and the dielectric constant in vacuum $= 1$. Note that frequently the zero component of a Hermitian biquaternion has retained another name (here $x_0 = t$) for historical reasons. Note also that

$$(ab)^* = a^* b^*, \ (ab)^\dagger = b^\dagger a^\dagger$$

for biquaternions.

A *Lorentz transformation* is to preserve the norm of $x = t + i\mathbf{x}$ and has the form

$$x \mapsto qxq^{*\dagger},$$

with

$$q = u + iv \ (u, v \in \mathbf{H})$$

satisfying $qq^\dagger = 1$, that is

$$uu^\dagger - vv^\dagger = 1, \ uv^\dagger + vu^\dagger = 0.$$

It is a *rotation* if $q^* = q$, that is $v = 0$, and a *boost* if $q^* = q^\dagger$, that is $u = u_0$ and $v_0 = 0$.

LEMMA 2.1. Every Lorentz transformation consists of a rotation followed by a boost.

Proof: If

$$\mu^2 = uu^\dagger = 1 + vv^\dagger \geq 1,$$

let

$$r = u\mu^{-1}, \ s = \mu - iuv^\dagger\mu^{-1},$$

then

$$sr = (\mu - iuv^\dagger\mu^{-1})u\mu^{-1} = u + iv.$$

Before turning to the intended physical application, let me say something about the *grammar* of the biquaternion language adapted to special relativity. Every significant physical quantity (or operator) is represented by a biquaternion together with an instruction how it is transformed under a co-ordinate transformation. In particular, many such quantities are represented by Hermitian biquaternions and transform like $x \mapsto qxq^{*\dagger}$. When multiplying two such Hermitian biquaternions a and b, we obtain

$$ab = qa\underbrace{q^{*\dagger}q}bq^{*\dagger}$$

and we are stuck with $q^{*\dagger}q$ in the middle. On the other hand, there is no problem with

$$ab^* \mapsto qa\underbrace{q^{*\dagger}q^*}b^*q^\dagger = qab^*q^\dagger,$$

even though it transforms differently.

In fact, so do its scalar and vector parts

$$S(ab^*) = \frac{1}{2}(ab^* + ba^*) = a_0 b_0 - \mathbf{a} \circ \mathbf{b}$$

and

$$V(ab^*) = \frac{1}{2}(ab^* - ba^*) = \mathbf{a} \times \mathbf{b} + i(\mathbf{a} b_0 - a_0 \mathbf{b}),$$

the latter being a "skew-Hermitian" biquaternion.

It will be convenient to introduce the *scalar product* of two Hermitian biquaternions:

$$a \odot b = a_0 b_0 - \mathbf{a} \circ \mathbf{b} = b \odot a.$$

This plays a rôle in the following useful

LEMMA 2.2. If a, b and c are Hermitian biquaternions, the Hermitian part of $ab^\dagger c$ is

$$a(b \odot c) - b(c \odot a) + c(a \odot b).$$

Proof: A simple calculation shows that the vector part of $\mathbf{abc}$ is $-\mathbf{a}(\mathbf{b} \circ \mathbf{c}) + \mathbf{b}(\mathbf{c} \circ \mathbf{a}) - \mathbf{c}(\mathbf{a} \circ \mathbf{b})$.

Now

$$\begin{aligned} ab^\dagger c = & \, a_0 b_0 c_0 + i(a_0 b_0 \mathbf{c} + b_0 c_0 \mathbf{a} - c_0 a_0 \mathbf{b}) \\ & + a_0 \mathbf{bc} - \mathbf{a} b_0 \mathbf{c} + \mathbf{ab} c_0 + i\mathbf{abc} \end{aligned}$$

From this we select the real scalar

$$a_0 b_0 c_0 - a_0(\mathbf{b} \circ \mathbf{c}) + b_0(\mathbf{c} \circ \mathbf{a}) - c_0(\mathbf{a} \circ \mathbf{b})$$

and the imaginary vector

$$i\{a_0 b_0 \mathbf{c} + b_0 c_0 \mathbf{a} - c_0 a_0 \mathbf{b} - \mathbf{a}(\mathbf{b} \circ \mathbf{c}) + \mathbf{b}(\mathbf{c} \circ \mathbf{a}) - \mathbf{c}(\mathbf{a} \circ \mathbf{b})\}.$$

Together these yield

$$a(b \odot c) - b(c \odot a) + c(a \odot b)$$

as was to be proved.[2)]

3. Special relativity and Maxwell's equations.

A physical quantity of interest is the *kinetic energy-momentum* biquaternion

$$p = p_0 + ip_0 \mathbf{v}, \ p_0 = E$$

where

$$\mathbf{v} = \frac{dx_1}{dt} i_1 + \frac{dx_2}{dt} i_2 + \frac{dx_3}{dt} i_3$$

is the classical velocity vector and E is the (kinetic) energy. If the rest-mass $m \neq 0$, we may put

$$p = m\frac{dx}{ds}, \ p_0 = m\frac{dt}{ds} = m(1 - v^2)^{-\frac{1}{2}}$$

where v is the absolute value of $\mathbf{v}$ and

$$ds^2 = N(dx) = dt^2 - dx_1^2 - dx_2^2 - dx_3^2.$$

Both p and dx/ds transform like $x \mapsto qxq^{*\dagger}$. According to Section 2, the product xp^* transforms as $xp^* \mapsto qxp^*q^\dagger$ and so do its scalar and vector parts

$$\frac{1}{2}(xp^* \pm px^*).$$

The latter is the relativistic *angular momentum*

$$\mathbf{x} \times \mathbf{p} + i(\mathbf{x}E - t\mathbf{p}),$$

where $\mathbf{x} \times \mathbf{p}$ is usually called the "angular momentum".

In the absence of an external force, $\frac{dp}{ds} = 0$, hence

$$\frac{d}{ds}(xp^*) = \frac{dx}{ds}p^* = \frac{dx}{ds}\left(\frac{dx}{ds}\right)^* m = m$$

and so

$$\frac{d}{ds}(x \odot p) = m \text{ and } \frac{d}{ds}V(xp^*) = 0.$$

The second equation asserts that

$$\frac{d}{ds}(\mathbf{x} \times \mathbf{p}) = 0 \text{ and } \frac{d}{ds}(\mathbf{x}E - t\mathbf{p}) = 0,$$

showing that the usual angular momentum is conserved, as well as the quantity $\mathbf{x}E - t\mathbf{p}$. (According to Penrose [loc.cit. p433], the latter conservation expresses the uniform motion of the center of mass.)

The differential operator

$$D = \frac{\partial}{\partial t} - i\nabla,\ \nabla = \frac{\partial}{\partial x_1}i_1 + \frac{\partial}{\partial x_2}i_2 + \frac{\partial}{\partial x_3}i_3,$$

is also a Hermitian biquaternion which transforms like x, namely $D \mapsto qDq^{*\dagger}$.

Maxwell's equations are usually presented in the Heaviside notation as follows:

$$\nabla \circ \mathbf{B} = 0,\ -\nabla \circ \mathbf{E} + \rho = 0,$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0,\ \frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{B} + \mathbf{J} = 0.$$

Combining the magnetic field $\mathbf{B}$ and the electric field $\mathbf{E}$ into a single biquaternion

$$F = \mathbf{B} + i\mathbf{E} \mapsto q^*Fq^{*\dagger}$$

and the charge density ρ with the current density $\mathbf{J}$ into the *charge-current density*

$$J = \rho + i\mathbf{J} \mapsto qJq^{*\dagger}$$

we may combine the four Maxwell equations into one:

$$DF + J = 0.$$

It follows that

$$D^*DF = -D^*J.$$

Here the scalar part of the left side is zero, hence so must be that of the right side. Thus we obtain the *equation of continuity*

$$\frac{\partial \rho}{\partial t} + \nabla \circ \mathbf{J} = 0.$$

The *charge-current* of an electrically charged particle is represented by the Hermitian biquaternion

$$e\frac{dx}{ds} = e\frac{dt}{ds}(1 + i\mathbf{v}),$$

where e is the charge. Multiplying this by $F = \mathbf{B} + i\mathbf{E}$ and putting

$$\frac{dt}{ds} = \gamma = (1 - v^2)^{-1/2},$$

we obtain

$$e\frac{dx}{ds}F = e\gamma(1 + i\mathbf{v})(\mathbf{B} + i\mathbf{E}).$$

The Hermitian part of this is

$$H\left(e\frac{dx}{ds}F\right) = e\gamma(\mathbf{v} \circ \mathbf{E} + i(\mathbf{E} + \mathbf{v} \times \mathbf{B}),$$

where $e(\mathbf{E} + \mathbf{v} \circ \mathbf{B})$ is usually called the *Lorentz force*. We thus have a relativistic version of the Lorentz force and are led to require that

$$\frac{dp}{ds} = H\left(e\frac{dx}{ds}F\right).$$

(We may think of $e\mathbf{v} \circ \mathbf{E}$ as the *Lorentz power*.)

As is well-known, Maxwell's equations imply the existence of a *four-potential*

$$A = \phi + i\mathbf{A}$$

such that

$$\frac{\partial \mathbf{A}}{\partial t} + \nabla\phi = -\mathbf{E}, \ \nabla \times \mathbf{A} = \mathbf{B}.$$

In other words, the vector part

$$V(D^*A) = -F. \tag{3.1}$$

However, A is not determined by this: we might equally well replace A by $A' = A - D\Theta$, where $\Theta = \Theta(x)$ and hence $D^*D\Theta$ are scalars. We will leave the choice of an appropriate Θ until later.

It is customary to require that

$$S(D^*A) = D \odot A = \frac{\partial \phi}{\partial t} + \nabla \circ \mathbf{A} = 0$$

so that (3.1) simplifies to

$$D^*A = -F$$

but we will have no need for this simplification.

A particle with electric charge e in an electro-magnetic field is said to have potential energy $e\phi$. But special relativity requires that this be accompanied by a potential momentum $e\mathbf{A}$, giving rise to relativistic Hermitian biquaternion

$$eA = e\phi + ie\mathbf{A}$$

representing the *potential energy-momentum.*

We should therefore expect the conservation of the total energy-momentum:

$$\frac{d}{ds}(p + eA) = 0.$$

It turns out that this is indeed the case, but only after A has been replaced by $A - D\Theta$ for an appropriate scalar field Θ.

Let us apply Lemma 2.2 to the triple product $(e\dot{x})D^*A$, where

$e\dot{x} = e\dfrac{dx}{ds} =$ charge-current,

$D^* = \dfrac{\partial}{\partial t} + i\nabla =$ conjugate partial differential operator,

$A = \phi + i\mathbf{A} =$ four-potential.

By Lemma 2.2, the Hermitian part

$$H(e\dot{x}D^*A) = e\dot{x}(D \odot A) - D(A \odot e\dot{x}) + A(e\dot{x} \odot D).$$

Now

$$H(e\dot{x}D^*A) - e\dot{x}(D \odot A) = H(e\dot{x}V(D^*A)) = -H(e\dot{x}F)$$

is the negative of the relativistic Lorentz force dp/ds and

$$(e\dot{x} \odot D)A = \frac{d}{ds}(eA)$$

hence

$$\frac{d}{ds}(p + eA) = D(A \odot e\dot{x}).$$

Having expected the right side to be zero, I was puzzled by the term $A \odot e\dot{x}$. This seems to be related to the quantum mechanical Bohm-Aharonov effect.[3)]

Putting $\dfrac{dx}{ds} \odot A = \dfrac{d\Theta}{ds}$, we have

$$D(A \odot e\dot{x}) = D\left(e\frac{d\Theta}{ds}\right) = \frac{d}{ds}(eD\Theta)$$

and so

$$\frac{d}{ds}(p + e(A - D\Theta)) = 0,$$

which suggests that we replace the original four-potential A by $A' = A - D\Theta$.

Repeating the same process for A' instead of A, we calculate

$$e\dot{x} \odot A' = e\dot{x} \odot A - e\dot{x} \odot D\Theta = 0.$$

To ensure that $A = A'$, we may postulate

$$e\dot{x} \odot A = 0,$$

in place of the usual assumption that $D \odot A = 0$, thus restricting the potential to its component orthogonal to the current in Minkowski space, and then we indeed have *conservation of the total energy momentum*:

$$\frac{d}{ds}(p + eA) = 0.$$

4. Quaternions and coquaternions.

Let us return to the division ring **H** of real quaternions. There are two so-called *regular representations* of **H** by linear transformations of the four-dimensional real vector space $\mathbf{R}^4$. The *left* representation $L(a)$ of the quaternion a sends the column vector $[x]$, made up from the coefficients of x, onto the column vector $[ax]$, while the *right* representation $R(a)$ sends $[x]$ onto $[xa]$. Thus

$$L(a)[x] = [ax], \;\; R(a)[x] = [xa].$$

It is easily seen that L preserves and R reverses composition:

$$L(aa') = L(a)L(a'), \;\; R(aa') = R(a')R(a),$$

and that

$$L(a)R(b) = R(b)L(a),$$

as follows from the associative law

$$a(xb) = (ax)b.$$

For example,

$$L(i_1)[x] = [i_1 x] = [-x_1 + i_1 x_0 - i_2 x_3 + i_3 x_2],$$

hence

$$L(i_1) = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = -e_{01} + e_{10} - e_{23} + e_{32}.$$

Here $e_{\alpha\beta}$ denotes the matrix with 1 in the intersection of column α and row β and 0 everywhere else, while α and β vary from 0 to 3.

It will be convenient to identify the quaternion a with the matrix $L(a)$. Then we have

$$i_1 = e_{01} - e_{10} - e_{23} + e_{32},$$

and i_2 and i_3 are obtained by cyclic permutation of the subscripts 1, 2 and 3.

Utilizing the diagonal matrix

$$\triangle = \text{ diag } (1, -1, -1, -1),$$

we see that

$$\triangle[x] = [x^\dagger],$$

hence

$$\begin{aligned} R(a)[x] &= [xa] = [(a^\dagger x^\dagger)^\dagger] \\ &= \triangle[a^\dagger x^\dagger] = \triangle L(a^\dagger)[x^\dagger] \\ &= \triangle L(a^\dagger)\triangle[x], \end{aligned}$$

so that

$$R(a) = \triangle L(a^\dagger)\triangle.$$

It will also be convenient to write

$$j_\alpha = R(i_\alpha),$$

then

$$j_0 = 1, \; j_1^2 = j_2^2 = j_3^2 = j_3 j_2 j_1 = -1.$$

We will say that

$$b_0 + b_1 j_1 + b_2 j_2 + b_3 j_3$$

is a *coquaternion.* The coquaternions form a division algebra $\mathbf{H}^{op}$. (Of course, $\mathbf{H}^{op} \simeq \mathbf{H}$ by conjugation.) One easily calculates

$$\begin{aligned} j_1 &= e_{01} - e_{10} - e_{23} + e_{32}, \\ i_1 j_1 &= e_{00} - e_{11} + e_{22} + e_{33}, \\ i_2 j_3 &= e_{01} + e_{10} - e_{23} - e_{32}, \\ i_3 j_2 &= -e_{01} - e_{10} - e_{23} - e_{32}, \end{aligned}$$

and obtains other such equations by cyclic permutation of 1, 2 and 3.

Adding and subtracting the equations which express the $i_\alpha j_\beta$ in terms of the $e_{\gamma\delta}$, we obtain

$$\begin{aligned} e_{00} &= \tfrac{1}{4}(i_0 j_0 - i_1 j_1 - i_2 j_2 - i_3 j_3), \\ e_{01} &= \tfrac{1}{4}(i_0 j_1 + i_1 j_0 + i_2 j_3 - i_3 j_2), \\ e_{11} &= \tfrac{1}{4}(i_0 j_0 - i_1 j_1 + i_2 j_2 + i_3 j_3), \\ e_{23} &= \tfrac{1}{4}(-i_0 j_1 + i_1 j_0 - i_2 j_3 - i_3 j_2). \end{aligned}$$

We note that $e_{\alpha\beta} = e_{\beta\alpha}^T$ is the *transposed* matrix of $e_{\beta\alpha}$ and that

$$\begin{aligned} i_\alpha^T &= i_\alpha^\dagger = -i_\alpha \text{ when } \alpha \neq 0, \\ j_\beta^T &= j_\beta^* = -j_\beta \text{ when } \beta \neq 0. \end{aligned}$$

Recall that * denotes the complex conjugate. Other equations are obtained by cyclic permutation of the subscripts 1, 2 and 3. Here and later, we denote the coquaternion conjugate by an asterisk.

It readily follows that every 4×4 real matrix can be written uniquely as

$$\begin{aligned} A &= \sum_{\alpha,\beta=0}^{3} j_\alpha i_\beta a_{\alpha\beta} \\ &= A_0 + j_1 A_1 + j_2 A_2 + j_3 A_3, \end{aligned}$$

when the A_α are quaternions, and

$$A = A'_0 + i_1 A'_1 + i_2 A'_2 + i_3 A'_3,$$

when the A'_β are coquaternions. Thus, given a quaternion x, we have

$$A[x] = [\sum i_\beta a_{\alpha\beta} x i_\alpha].$$

The matrix A is

diagonal iff $A = \sum j_\alpha i_\alpha a_{\alpha\alpha}$,
symmetric iff $a_{\alpha\beta} = 0$ when $\alpha = 0$ and $\beta = 0$ or both $\neq 0$,
skew-symmetric iff $a_{\alpha\beta} = 0$ unless $\alpha = 0$ and $\beta = 0$ or both $\neq 0$.

Note that the *transposed* matrix A^T is given by

$$\begin{aligned} A^T = A^{*\dagger} &= A_0^\dagger - j_1 A_1^\dagger - j_2 A_2^\dagger - j_3 A_3^\dagger \\ &= A_0'^* - i_1 A_1'^* - i_2 A_2'^* - i_3 A_3'^*. \end{aligned}$$

To put the above observations into a wider algebraic context: if $\mathbf{H}$ is the division algebra of quaternions, any $\mathbf{H} - \mathbf{H}$-bimodule may be viewed as a left $\mathbf{H}^{op} \otimes \mathbf{H}$-module. In view of the well-known fact that $\mathbf{H}^{op} \otimes \mathbf{H}$ is isomorphic to the ring of all 4×4 real matrices, any linear transformation of a four-dimensional real vector space can be expressed with the help of pre- and post-multiplication of quaternions by quaternions. This observation has been made repeatedly by linear algebraists as well as by physicists. It is a special case of a well-known theorem about *central simple algebras* (see e.g. Jacobson [1980]).

5. Quaternions and the Dirac equation.

So far, we have looked at the application of quaternions to special relativity. Among applications to quantum mechanics, I will single out the celebrated Dirac equation for the electron. Dirac himself originally derived this using what is now seen as an argument involving Clifford algebras, but he repeatedly expressed the hope that this could also be done with the help of quaternions, though not with biquaternions.

The reason why biquaternions no longer suffice is that we require entities which anticommute with what we called i. We are then led to replace the imaginary number i by a suitable coquaternion, for example by $j_1 = R(i_1)$ or, more generally, by $R(ri_1r^\dagger)$, where r is a real quaternion of norm 1. This involves a paradigm shift from biquaternions to 4×4 real matrices. What we had called Hermitian biquaternions now become symmetric matrices

$$x = x_0 + j_1 i_1 x_1 + j_1 i_2 x_2 + j_1 i_3 x_3.$$

This raises a question: what happens to the missing six special dimensions generated by $j_2 i_\alpha$ and $j_3 i_\alpha$ with $\alpha = 1, 2$ or 3? Curiously, had we represented position in space-time by the skew-symmetric matrix

$$j_1 x_0 + i_1 x_1 + i_2 x_2 + i_3 x_3$$

instead, we would be looking for two missing time dimensions generated by j_2 and j_3 instead.

The second order wave equation

$$DD^*\Psi = -m^2\Psi,$$

where $\Psi = \Psi(x)$, is known as the *Klein-Gordon* equation, although Schrödinger was certainly familiar with it before turning to his non-relativistic formulation. Here

$$DD^* = \frac{\partial}{\partial t^2} - \nabla \circ \nabla$$

and m is the rest-mass of the electron (or any fermion for that matter), a scalar constant. For the moment, $\Psi = \Psi(x)$ is any (smooth) real matrix function of x.

PROPOSITION 5.1. The second order Klein-Gordon equation is equivalent to the first order Dirac equation

$$D^*\Psi = -j_2 m\Psi$$

provided m is invertible.

Proof. To derive the Klein-Gordon equation from the Dirac equation we need only observe that $Dj_2 = j_2 D^*$, since j_2 anticommutes with j_1. To go in the other direction, suppose $\Psi = \Psi'$ is a solution of the Klein-Gordon equation. Putting

$$D^*\Psi' = m\Psi'',$$

we obtain

$$D\Psi'' = -m\Psi'.$$

Now letting

$$\Psi = \Psi' + j_2\Psi''$$

and recalling that j_2 anticommutes with j_1, one readily obtains the Dirac equation.

The Dirac equation for the free electron is usually written thus:

$$i\left(\gamma_0\frac{\partial}{\partial t} - \gamma_1\frac{\partial}{\partial x_1} - \gamma_2\frac{\partial}{\partial x_2} - \gamma_3\frac{\partial}{\partial x_3}\right)\Psi \equiv m\Psi,$$

when the $i\gamma_\alpha$ are specially constructed complex matrices due to Pauli. I prefer the real matrices $j_\alpha = R(i_\alpha)$, which seem more natural.

The above derivation yields an explicit solution of the Dirac equation if we begin with the obvious solution

$$\Psi' = \cos(x \odot p)\Psi_0$$

of the Klein-Gordon equation, Ψ_0 being a constant matrix. For then we have

$$\Psi'' = \frac{1}{m}D^\dagger \cos(x \odot p)\Psi_0 = -\frac{1}{m}p^\dagger \sin(x \odot p)\Psi_0,$$

hence

$$\Psi = (\cos(x \odot p) + \eta \sin(x \odot p))\Psi_0$$

where $\eta = -\dot{x}j_2$. Exploiting the fact that

$$\eta^2 = (\dot{x}j_2)(j_2\dot{x}^\dagger) = -\dot{x}\dot{x}^\dagger = -1,$$

we may write this more elegantly as

$$\Psi = \exp(\eta(x \odot p))\Psi_0.$$

Our argument applies to any particle of non-zero rest-mass, and this nowadays includes the neutrino. But how does Ψ behave under a Lorentz transformation? For fermions of spin $\frac{1}{2}$, Dirac insisted that Ψ was a *spinor*. In our language this would mean that

$$\Psi \mapsto q\Psi.$$

The real matrix Ψ may be multiplied by any constant column vector $[c]$ on the right, giving rise to the column vector $\Psi[c] = [\psi_c]$, where ψ_c is a real quaternion. Hence the Dirac equation may be written

$$\left(\frac{\partial}{\partial t} + j_1\nabla\right)[\psi_c] = -j_2[m\psi_c],$$

that is,

$$\left[\frac{\partial}{\partial t}\psi_c + \nabla\psi_c i_1\right] \;=\; [-m\psi_c i_2]$$

that is,

$$\frac{\partial}{\partial t}\psi_c + \nabla\psi_c i_1 = -m\psi_c i_2, \tag{5.2}$$

involving quaternions only. If we replace the i_α on the right by $ri_\alpha r^\dagger$, r being a quaternion of norm 1, we obtain

$$\frac{\partial}{\partial t}\psi_c r + \nabla\psi_c r i_1 = -m\psi_c r i_2.$$

This amounts to replacing ψ_c by $\psi_c r = \psi_{cr}$. (I don't know whether any physical significance can be attached to the choice of r.)

6. Epilogue.

Having been introduced to quaternions by my teacher Gordon Pall in an undergraduate number theory course 65 years ago, I found the book by Silberstein [1924] in the library, which offered me a handle for grasping the theory of special relativity and tempted me to approach the Dirac equation. However, Dirac himself discouraged me, since I was still attached to biquaternions. At the same time, I learned from Arthur Conway how to get around this problem, but his impressive paper [1948] persuaded me that he had already done all I was planning to do. (His paper even went beyond what is being presented here: in a final section it derives the hydrogen line spectrum and its fine structure.) I had met both Dirac and Conway at a summer school in Vancouver in 1949, and it may have been the former's criticism and the latter's achievement that persuaded me to abandon physics for mathematics.

I confess I have not made a thorough study of the literature. The references below list only those publications that have come my way in one way or another and include a number of popular expositions, in particular the marvellous semi-popular discussion by Feynman [1985]. Adler [1995] has a more extensive list of references, but with only a small intersection with mine. I apologize to those authors whose contributions I have missed.

As far as I can tell, the main results on the application of quaternions presented here are due to A.W. Conway [1948] and F. Gürsey [1955,1958] ignoring the even earlier contributions by Cornelius Lanczos [1929], which came to my attention only quite recently. See the insightful critical discussion by A. Gsponer and J.-P. Hurni [1998], which puts the Lanczos work into its historical context and offers some interesting further speculation.

I doubt whether physicists will find anything new in the above exposition, but teachers of physics may be interested in the description of the Lorentz force[4)] in Section 3 as $\frac{d}{ds}(eA)$, provided the four-potential A has been subjected to a gauge transformation ensuring that A is orthogonal to $\dot{x}$, and to the explicit solution of the Dirac equation in Section 5, which I have not found in any of the texts I consulted.

Endnotes

[1)] Some authors have speculated about the middle name “Olinde”, apparently unaware that this was the name of a former Dutch colony, now the old town of Recife in Brazil, which may throw some light on the history of the Rodrigues family, perhaps on the maternal side.

[2)] A search of the literature reveals that this identity first appears in Gürsey [1955].

[3)] I am tempted to conjecture that it serves to eliminate the force which a moving particle exerts on itself.

[4)] I take this opportunity to point out that the description of the Lorentz force in my 1995 article is wrong and that the complex matrices there have been replaced by real matrices here.

Acknowledgement

I wish to thank the anonymous referee for his careful reading of the manuscript and his constructive criticism.

References

[1] S.L. Adler, Composite leptons and quarks constructed as triply occupied quasiparticles in quaternionic quantum mechanics, *Physics Letters*. B 332 (1994), 358-365.

[2] S.L. Adler, *Quaternions, quantum mechanics and quantum fields*. Oxford University Press 1995.

[3] S.L. Altmann, *Rotations, quaternions and double groups*. Clarendon Press, Oxford 1986.

[4] S.L. Altmann, Hamilton, Rodrigues and the quaternion scandal, *Math. Magazine* 62 (1989), 291-308.

[5] S.P. Brumby and G.C. Joshi, *Experimental Status of Quaternionic Quantum Mechanics*, University of Melbourne Report UM-P-94/120.

[6] A.W. Conway, On the application of quaternions to some recent developments of electrical theory, *Proc. Royal Irish Acad.* A29(1911), 80.

[7] A.W. Conway, The quaternionic form of relativity, *Phil. Mag.* 24(1912), 208.

[8] A.W. Conway, Quaternions and quantum mechanics, *Pontificia Academia Scientiarum* 12(1948), 204-277.

[9] A.W. Conway, His life work and influence, *Proc. Second Canadian Math. Congress*, Toronto: University of Toronto Press (1951).

[10] W.H. Davis et al. (eds), *Cornelius Lanczos' collected papers with commentaries*, Volume iii, North Carolina State University, Raleigh 1998.

[11] P.A.M. Dirac, Applications of quaternions to Lorentz transformations, *Proc. Roy. Irish Acad.* A50(1945), 261-270.

[12] R.P. Feynman, *QED, the strange theory of light and matter*, Princeton University Press, Princeton NJ 1985.

[13] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman Lectures on Physics*. Addison Wesley Publ. Co., Reading Mass. 1963/1977.

[14] D. Finkelstein, J.M. Jauch and D. Speiser, Notes on quantum mechanics, in: Hooker (ed.), *Logico-algebraic approach to quantum mechanics* II, 367-421, Reidel Publ. Co., Dordrecht 1979.

[15] M. Gell-Mann, *The quark and the jaguar*, W.H. Freeman, New York 1994.

[16] A. Gsponer and J.-P. Hurni, *Lanczos-Einstein-Petiau: from Dirac's equation to non-linear wave mechanics*, in: W.H. Davis et al, [1998].

[17] F. Gürsey. Contributions to the quaternionic formalism in special relativity, *Rev. Faculté Sci. Univ. Istanbul* A20(1955), 149-171.

[18] F. Gürsey, Correspondence between quaternions and four-spinors, *Rev. Faculté Sci. Univ. Istanbul* A21(1958), 33-54.

[19] N. Jacobson, *Basic Algebra* II, Freeman, San Francisco 1980.

[20] Yuntong Kuo, *A discussion of quaternions*, Manuscript 1996.

[21] J. Lambek, *Biquaternion vector fields over Minkowski space*. Thesis, Part I, McGill University, Montreal 1950.

[22] J. Lambek, If Hamilton had prevailed; quaternions in physics, *The mathematical intelligencer* 17(1995), 7-15.

[23] J. Lambek, *Quaternions and the particles of nature*, Dept. of Mathematics, McGill University Report 1996-03.

[24] C. Lanczos, Three articles on Dirac's equation, Zeitschrift für Physik 57(1929), 447-493 (reprinted in W.H. Davis et al. [1998]).

[25] N. Mackey, Hamilton and Jacobi meet again: quaternions and the eigenvalue problem, *SIAM J. Matrix Anal. Appl.* 18(1995), 421-435.

[26] J.C. Maxwell, Remarks on the mathematical classification of physical quantities, *Proc. London Math. Soc.* 3(1869), 224-232.

[27] W. Moore, *Schrödinger, Life and Thought*, Cambridge, Cambridge University Press (1989).

[28] P.J. Nahin, Oliver Heaviside, *Sci. American* 1990, 122-129.

[29] R. Penrose, *The road to reality: a complete guide to the laws of the universe*, Jonathan Cape, London 2004.

[30] L.H. Ryder, *Quantum field theory*, Cambridge University Press 1985/1997.

[31] L. Silberstein, Quaternionic form of relativity, *Phil. Mag.* 23(1912), 790.

[32] L. Silberstein, *Theory of relativity*, MacMillan, London 1924.

[33] A. Sudbery, *Quantum mechanics and the particles of nature*. Cambridge University Press, Cambridge 1986.

[34] J.L. Synge, Quaternions, Lorentz transformations and the Conway-Dirac-Eddington matrices, *Commun. Dublin Inst. Adv. Studies* A21(1972), 1-67.

[35] V. Trifonov, Natural geometry and nonzero quaternions, *Int. J. of Theoretical Physics 46* (2) (2007), 251-257.

[36] J. von Neumann and G. Birkhoff, The logic of quantum mechanics, *Annals of Math.* 37(1936), 823-843.

[37] P. Weiss, On some applications of quaternions to restricted relativity and classical radiation theory, *Proc. Roy. Irish Acad.* A46(1941), 129-168.

[38] H. Weyl, *The theory of groups and quantum mechanics*, Dover Publication, New York (1950).

Appendix

This is an exposition of the algebra of biquaternions that I wrote to help me understand better what Jim was doing. I was especially interested to understand the function of the two involutions whose composite was matrix transposition. I have not tried to coordinate my notation with his.

The real algebra $\mathbf{H} \otimes \mathbf{H}^{\mathrm{op}} \cong \mathbf{H} \otimes \mathbf{H} \cong \mathrm{Mat}_4(\mathbf{R})$ (from the theory of the Brauer group) can be described as follows. We define 2×2 matrices $\mathbf{r}, \mathbf{s}, \mathbf{q}$ which, together with the 2×2 identity we will denote by $\mathbf{1}$, allows us to define the tensor product using 2×2 block matrices $\mathbf{1}, \mathbf{r}, \mathbf{s}, \mathbf{q}$ as blocks.

The building blocks are the 2×2 matrices

$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (identity),

$\mathbf{r} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ (rotation around the x-axis),

$\mathbf{s} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (rotation around the line $x = y$), and

$\mathbf{q} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (quarter turn counter-clockwise).

We denote by $\cdot$ the dot product of $n \times n$ matrices as vectors in n^2 dimensional real space.

0.1. Proposition. *We have that*

$$\begin{array}{lll} \mathbf{r}^2 = \mathbf{1} & \mathbf{s}^2 = \mathbf{1} & \mathbf{q}^2 = -\mathbf{1} \\ \mathbf{rs} = -\mathbf{q} = -\mathbf{sr} & \mathbf{qr} = \mathbf{s} = -\mathbf{rq} & \mathbf{sq} = \mathbf{r} = -\mathbf{qs} \end{array}$$

$$\mathbf{1} \cdot \mathbf{1} = \mathbf{r} \cdot \mathbf{r} = \mathbf{s} \cdot \mathbf{s} = \mathbf{q} \cdot \mathbf{q} = 2$$

$$\mathbf{1} \cdot \mathbf{r} = \mathbf{1} \cdot \mathbf{s} = \mathbf{1} \cdot \mathbf{r} = \mathbf{1} \cdot \mathbf{q} = \mathbf{r} \cdot \mathbf{s} = \mathbf{r} \cdot \mathbf{q} = \mathbf{s} \cdot \mathbf{q} = 0$$

These can be shown by direct computation.

We let (using block matrices, so these are 4×4)

$$\begin{array}{llll} \mathbf{i}_1 = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} & \mathbf{i}_2 = \begin{pmatrix} \mathbf{q} & \mathbf{0} \\ \mathbf{0} & \mathbf{q} \end{pmatrix} & \mathbf{i}_3 = \begin{pmatrix} \mathbf{0} & -\mathbf{r} \\ \mathbf{r} & \mathbf{0} \end{pmatrix} & \mathbf{i}_4 = \begin{pmatrix} \mathbf{0} & -\mathbf{s} \\ \mathbf{s} & \mathbf{0} \end{pmatrix} \\ \mathbf{j}_1 = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} & \mathbf{j}_2 = \begin{pmatrix} \mathbf{q} & \mathbf{0} \\ \mathbf{0} & -\mathbf{q} \end{pmatrix} & \mathbf{j}_3 = \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} & \mathbf{j}_4 = \begin{pmatrix} \mathbf{0} & \mathbf{q} \\ \mathbf{q} & \mathbf{0} \end{pmatrix} \end{array}$$

In the following, the integer variables x and y range over $[1, 4]$.

The matrices $\mathbf{i}_x$ represent the left regular representation of $\mathbf{H}$ on itself and the matrices $\mathbf{j}_x$ represent the right regular representation. As a result the $\mathbf{i}_x$ are isomophic to $\mathbf{H}$, while the $\mathbf{j}_x$ represent $\mathbf{H}^{\mathrm{op}}$ and for all x, y, $\mathbf{i}_x$ commutes with $\mathbf{j}_y$. Now let $\mathbf{a}_{xy} = \mathbf{i}_x \mathbf{j}_y$.

0.2. Proposition. *We have*

$$\mathsf{a}_{11} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \quad \mathsf{a}_{12} = \begin{pmatrix} \mathsf{q} & \mathbf{0} \\ \mathbf{0} & -\mathsf{q} \end{pmatrix} \quad \mathsf{a}_{13} = \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{14} = \begin{pmatrix} \mathbf{0} & \mathsf{q} \\ \mathsf{q} & \mathbf{0} \end{pmatrix}$$

$$\mathsf{a}_{21} = \begin{pmatrix} \mathsf{q} & \mathbf{0} \\ \mathbf{0} & \mathsf{q} \end{pmatrix} \quad \mathsf{a}_{22} = \begin{pmatrix} -\mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \quad \mathsf{a}_{23} = \begin{pmatrix} \mathbf{0} & -\mathsf{q} \\ \mathsf{q} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{24} = \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}$$

$$\mathsf{a}_{31} = \begin{pmatrix} \mathbf{0} & -\mathsf{r} \\ \mathsf{r} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{32} = \begin{pmatrix} \mathbf{0} & -\mathsf{s} \\ -\mathsf{s} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{33} = \begin{pmatrix} -\mathsf{r} & \mathbf{0} \\ \mathbf{0} & -\mathsf{r} \end{pmatrix} \quad \mathsf{a}_{34} = \begin{pmatrix} \mathsf{s} & \mathbf{0} \\ \mathbf{0} & -\mathsf{s} \end{pmatrix}$$

$$\mathsf{a}_{41} = \begin{pmatrix} \mathbf{0} & -\mathsf{s} \\ \mathsf{s} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{42} = \begin{pmatrix} \mathbf{0} & \mathsf{r} \\ \mathsf{r} & \mathbf{0} \end{pmatrix} \quad \mathsf{a}_{43} = \begin{pmatrix} -\mathsf{s} & \mathbf{0} \\ \mathbf{0} & -\mathsf{s} \end{pmatrix} \quad \mathsf{a}_{44} = \begin{pmatrix} -\mathsf{r} & \mathbf{0} \\ \mathbf{0} & \mathsf{r} \end{pmatrix}$$

0.3. Proposition. $\mathsf{a}_{xy} \cdot \mathsf{a}_{uv} = 4\delta_{xu}\delta_{yv}$*; hence the matrices* $\mathsf{a}_{xy}/2$ *are an orthomormal basis for* $\mathrm{Mat}_4(\mathsf{R})$.

Now define two operators $^\sharp$ and $^\flat$ on this space as the unique linear extension of the function defined on the basis by

$$\mathsf{a}_{xy}^\sharp = \begin{cases} \mathsf{a}_{xy} & \text{if } x = 1 \\ -\mathsf{a}_{xy} & \text{if } x = 2, 3, 4 \end{cases}$$

$$\mathsf{a}_{xy}^\flat = \begin{cases} \mathsf{a}_{xy} & \text{if } y = 1 \\ -\mathsf{a}_{xy} & \text{if } y = 2, 3, 4 \end{cases}$$

What this amounts to applying quaternionic conjugation to the i_x for $^\sharp$ and to the j_y for $^\flat$.

0.4. Proposition. *For any matrix* b, $(\mathsf{b}^\sharp)^\flat = (\mathsf{b}^\flat)^\sharp = \mathsf{b}^t$, *the transpose of* b.

This is not hard to prove by showing it is true for each a_{xy} and using the fact that all three operators are linear.

So we conclude that there are two involutory operations on $\mathrm{Mat}_4(\mathsf{R})$ whose composition is transpose. Is there any other dimension in which this happens? The above cannot be duplicated in other dimensions because, among other reasons, the fact that the right and left regular representations makes the orthogonality relations impossible.

Here is a formula for the operation, but it is sufficiently complicated, unlike the claims above, to preclude hand computation. Since the $\mathsf{a}_{xy}/2$ form an orthonormal basis, it is the case that for any $\mathsf{b} \in \mathrm{Mat}_4(\mathsf{R})$

$$\mathsf{b} = \frac{1}{4} \sum_{i=1,j=1}^{4,4} (\mathsf{a}_{xy} \cdot \mathsf{b})\mathsf{a}_{xy}$$

and then

$$\mathsf{b}^\sharp = \frac{1}{4} \sum_{i=1,j=1}^{4,4} (\mathsf{a}_{xy} \cdot \mathsf{b})\mathsf{a}_{xy}^\sharp$$

$$\mathsf{b}^\flat = \frac{1}{4} \sum_{i=1,j=1}^{4,4} (\mathsf{a}_{xy} \cdot \mathsf{b})\mathsf{a}_{xy}^\flat$$

Someone with computer mojo than us can perhaps find an explicit formula for these operations. Standard linear algebra packages do not implement dot product of matrices as far as we know.

We note that these formulas, although derived from considerations of $\mathbf{H} \otimes \mathbf{H}^{\mathrm{op}}$, are well defined over any field of characteristic different from 2 and even any commutative ring that is 2-divisible.

We further note that $\mathbf{a}_{12}, \mathbf{a}_{13}, \mathbf{a}_{14}, \mathbf{a}_{21}, \mathbf{a}_{31}, \mathbf{a}_{41}$ are all skew symmetric and, since the space of 4×4 skew symmetric matrices is 6-dimensional, these 6 matrices are a basis for that space.

Example: Let $\mathbf{e}_{41}$ be the matrix with a 1 in the first column, fourth row. Then we can calculate that $\mathbf{e}_{41} = \mathbf{a}_{14} + \mathbf{a}_{23} - \mathbf{a}_(32) + \mathbf{a}_{41}$, whence $\mathbf{e}_{41}^{\sharp} = \mathbf{a}_{14} - \mathbf{a}_{23} + \mathbf{a}_{32} - \mathbf{a}_{41}$. Also $\mathbf{e}_{41}^{t} = -\mathbf{a}_{14} + \mathbf{a}_{23} - \mathbf{a}_{32} - \mathbf{a}_{41}$, whence $\mathbf{e}_{14}^{\flat} = \mathbf{a}_{14} - \mathbf{a}_{23} + \mathbf{a}_{32} - \mathbf{a}_{41}$ which is exactly the same. You can calculate that

$$\mathbf{e}_{41}^{\sharp} = \mathbf{e}_{41}^{t\flat} = 1/2 \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

We do one more computation: let $\mathbf{e}_{11}$ be the matrix with a 1 in the first column first row. Then we can calculate that $\mathbf{e}_{11} = \mathbf{a}_{11} + \mathbf{a}_{22} - \mathbf{a}_{33} - \mathbf{a}_{44}$, whence $\mathbf{e}_{11}^{\sharp} = \mathbf{a}_{11} + \mathbf{a}_{22} + \mathbf{a}_{33} + \mathbf{a}_{44}$, which works out to

$$\mathbf{e}_{11}^{\sharp} = \mathbf{e}_{11}^{\flat} = 1/2 \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

With 14 more computations of this sort, we could write explicit formulas for $^{\sharp}$ and $^{\flat}$. This does not appear to be simple to program since matrix algebra programs do not generally implement dot product of matrices.

Is there any dimension n other than 4 for which there is a non-identity automorphism $^{\sharp}$ and an antiautomorphism $^{\flat}$ of the matrix ring $\mathrm{Mat}_n(\mathbf{R})$ which commute and whose composite is transpose?

$$\mathbf{e}_{23} = -\mathbf{a}_{14} - \mathbf{a}_{23} - \mathbf{a}_{32} + \mathbf{a}_{41}$$

$$\mathbf{e}_{23}^{\sharp} = 1/2 \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{e}_{12} = \mathbf{a}_{12} + \mathbf{a}_{21} + \mathbf{a}_{34} - \mathbf{a}_{43}$$

$$\mathbf{e}_{12}^{\sharp} = 1/2 \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

My best guess: First for $\mathbf{e}_{tt}$, let the other three indices be u, v, w, then $\mathbf{e}_{tt}^{\sharp} = \mathbf{e}_{tt}\flat = -\mathbf{e}_{tt} + \mathbf{e}_{uu} + \mathbf{e}_{vv} + \mathbf{e}_{ww}$. For $\mathbf{e}_{tu}$ with $t < u$, let the four indices be denoted t, u, v, w with $v < w$. Then $\mathbf{e}_{tu}^{*} = -\mathbf{e}_{tu} - \mathbf{e}_{vw} + \mathbf{e}_{wv} - \mathbf{e}_{ut}$. Finally, for $\mathbf{e}_{tu}$ with $t > u$, let the four indices be t, u, v, w with $v > w$ and use the same formula.

To put it in other terms, $\mathbf{e}_{tu}^{\sharp}$ is the sum of four terms, three with a - sign. The one + sign is the $\mathbf{e}_{vw}$ that is on the opposite side of the diagonal. The computation of $\mathbf{e}_{tu}\flat$ differs only in the + sign is on the same side of the diagonal.

0.5. A PARTIAL EXPLANATION of why the composite is an anti-automorphism.

0.6. PROPOSITION. *Suppose we have an algebra A, two subalgebras B and C of A, and two linear automorhisms σ and τ of A with the following properties for $b \in B$ and $c \in C$:*

- $A = BC$;
- $bc = cb$;
- $\sigma|B$ *and* $\tau|C$ *are anti-automorphisms of the algebra structures.;*
- $\sigma(bc) = \sigma(b)c$ *and* $\tau(bc) = b\tau(c)$

Then $\sigma\tau$ is an anti-automorphism of the algebra structure.

The proof is trivial. To apply it here, let A be the matrix algebra, B the subalgebra generated by $\mathbf{i}_x$, $x = 1, 2, 3, 4$ and C be the subalgebra generated by $\mathbf{j}_y$, $y = 1, 2, 3, 4$. Of course, $\sigma = {}^\sharp$ and $\tau = {}^\flat$

Department of mathematics and statistics
McGill University, Montreal