Doing Physics with Quaternions

Douglas B. Sweetser

INDEX

What Are Quaternions?

Quaternions are numbers like the real numbers: they can be added, subtracted, multiplied, and divided. There is something odd about them, hinted at by the Latin. Quaternions are composed of four numbers that work together as one. They were discovered by several people back in the eighteen hundreds. Some enthusiasts thought quaternions would be able to express everything that could happen in our three dimensions of space and one for time because quaternions naturally had that form too. Math accidents do not happen − they revel deep things about how Nature works.

Unfortunately, the big fans of quaternion mathematics claimed far more than they would deliver. Useful ideas born from initial quaternion work − for example the notion of scalars, vectors, div, grad, and curl − were stripped out of their initial context, and made more "general". I use the quotes because from my viewpoint, there is nothing more general than a number that can be added, subtracted, multiplied, and divided. I am trying to continue the project of applying four−dimensional quaternions to the four−dimensional spacetime we live in.

Unifying Two Views of Events

An experimentalist collects events about a physical system. A theorists builds a model to describe what patterns of events within a system might generate the experimentalist's data set. With hard work and luck, the two will agree!

Events are handled mathematically as 4−vectors. They can be added or subtracted from another, or multiplied by a scalar. Nothing else can be done. A theorist can import very powerful tools to generate patterns, like metrics and group theory. Theorists in physics have been able to construct the most accurate models of Nature in all of science.

I hope to bring the full power of mathematics down to the level of the events themselves. This may be done by representing events as the mathematical field of quaternions. All the standard tools for creating mathematical patterns − multiplication, trigonometric functions, transcendental functions, infinite series, the special functions of physics − should be available for quaternions. Now a theorist can create patterns of events with events. This may lead to a better unification between the work of a theorist and the work of an experimentalist.

An Overview of Doing Physics with Quaternions

It has been said that one reason physics succeeds is because all the terms in an equation are tensors of the same rank. This work challenges that assumption, proposing instead an integrated set of equations which are all based on the same 4−dimensional mathematical field of quaternions. Mostly this document shows in cookbook style how quaternion equations are equivalent to approaches already in use. As Feynman pointed out, "whatever we are *allowed* to imagine in science must be *consistent with everything else we know*." Fresh perspectives arise because, in essence, tensors of different rank can mix within the same equation. The four Maxwell equations become one nonhomogeneous quaternion wave equation, and the Klein−Gordon equation is part of a quaternion simple harmonic oscillator. Even gravity may be part of a simple quaternion wave equation, a research topic of much interest to me. Since all of the tools used are woven from the same mathematical fabric, the interrelationships become more clear. Hope you enjoy.

A Brief History of Quaternions

Complex numbers were a hot subject for research in the eighteen hundreds. An obvious question was that if a rule for multiplying two numbers together was known, what about multiplying three numbers? For over a decade, this simple question had bothered Hamilton, the big mathematician of his day. The pressure to find a solution was not merely from within. Hamilton wrote to his son:

"Every morning in the early part of the above−cited month [Oct. 1843] on my coming down to breakfast, your brother William Edwin and yourself used to ask me, ’Well, Papa, can you multiply triplets?’ Whereto I was always obliged to reply, with a sad shake of the head, ’No, I can only add and subtract them.’"

We can guess how Hollywood would handle the Brougham Bridge scene in Dublin. Strolling along the Royal Canal with Mrs. H−, he realizes the solution to the problem, jots it down in a notebook. So excited, he took out a knife and carved the answer in the stone of the bridge.

Hamilton had found a long sought−after solution, but it was weird, very weird, it was 4D. One of the first things Hamilton did was get rid of the fourth dimension, setting it equal to zero, and calling the result a "proper quaternion." He spent the rest of his life trying to find a use for quaternions. By the end of the nineteenth century, quaternions were viewed as an oversold novelty.

In the early years of this century, Prof. Gibbs of Yale found a use for proper quaternions by reducing the extra fluid surrounding Hamilton’s work and adding key ingredients from Rodrigues concerning the application to the rotation of spheres. He ended up with the vector dot product and cross product we know today. This was a useful and potent brew. Our investment in vectors is enormous, eclipsing their place of birth (Harvard had >1000 references under "vector", about 20 under "quaternions", most of those written before the turn of the century).

In the early years of this century, Albert Einstein found a use for four dimensions. In order to make the speed of light constant for all inertial observers, space and time had to be united. Here was a topic tailor−made for a 4D tool, but Albert was not a math buff, and built a machine that worked from locally available parts. We can say now that Einstein discovered Minkowski spacetime and the Lorentz transformation, the tools required to solve problems in special relativity.

Today, quaternions are of interest to historians of mathematics. Vector analysis performs the daily mathematical routine that could also be done with quaternions. I personally think that there may be 4D roads in physics that can be efficiently traveled only by quaternions, and that is the path which is laid out in these web pages.

In a longer history, Gauss would get the credit for seeing quaternions first in one of his notebooks. Rodrigues developed 3D rotations all on his own also in the 1840’s. The Pauli spin matrices and Penrose’s spinors are reinventions of the wheel that miss out on division. Although I believe that is a major omission and cause of subtle flaws at the foundations of modern physics, spin matrices and spinors have many more adherents today than quaternions.

Mathematics

Multiplying Quaternions the Easy Way

Multiplying two complex numbers a + b I and c + d I is straightforward.

$$(a, b)\,(c, d) = (ac - bd,\ ad + bc)$$

For two quaternions, b I and d I become the 3−vectors B and D, where B = x I + y J + z K and similarly for D. Multiplication of quaternions is like complex numbers, but with the addition of the cross product.

$$(a, \vec{B})\,(c, \vec{D}) = (ac - \vec{B}.\vec{D},\ a\,\vec{D} + \vec{B}\,c + \vec{B} \times \vec{D})$$

A mnemonic: (firsts − lasts, outside + inside + the cross). Note that the cross product would change its sign if the order of multiplication were reversed, unlike the other terms. That is why quaternions in general do not commute.

If a is the operator d/dt, and B is the del operator, or d/dx I + d/dy J + d/dz K (all partial derivatives), then these operators act on the scalar function c and the 3−vector function D in the following manner:

$$\left(\frac{d}{dt}, \vec{\nabla}\right)(c, \vec{D}) = \left(\frac{dc}{dt} - \vec{\nabla}.\vec{D},\ \frac{d\,\vec{D}}{dt} + \vec{\nabla}\,c + \vec{\nabla} \times \vec{D}\right)$$

This one quaternion contains the time derivatives of the scalar and 3−vector functions, along with the divergence, the gradient and the curl. Dense notation :−)

Scalars, Vectors, Tensors and All That

According to my math dictionary, a tensor is ...

"An abstract object having a definitely specified system of components in every coordinate system under consideration and such that, under transformation of coordinates, the components of the object undergoes a transformation of a certain nature."

To make this introduction less abstract, I will confine the discussion to the simplest tensors under rotational transformations. A rank−0 tensor is known as a scalar. It does not change at all under a rotation. It contains exactly one number, never more or less. There is a zero index for a scalar. A rank−1 tensor is a vector. A vector does change under rotation. Vectors have one index which can run from 1 to the number of dimensions of the field, so there is no way to know a priori how many numbers (or operators, or ...) are in a vector. n−rank tensors have n indices. The number of numbers needed is the number of dimensions in the vector space raised by the rank. Symmetry can often simplify the number of numbers actually needed to describe a tensor.

There are a variety of important spin−offs of a standard vector. Dual vectors, when multiplied by its corresponding vector, generate a real number, by systematically multiplying each component from the dual vector and the vector together and summing the total. If the space a vector lives in is shrunk, a contravariant vector shrinks, but a covariant vector gets larger. A tangent vector is, well, tangent to a vector function.

Physics equations involve tensors of the same rank. There are scalar equations, polar vector equations, axial vector equations, and equations for higher rank tensors. Since the same rank tensors are on both sides, the identity is preserved under a rotational transformation. One could decide to arbitrarily combine tensor equations of different rank, and they would still be valid under the transformation.

There are ways to switch ranks. If there are two vectors and one wants a result that is a scalar, that requires the intervention of a metric to broker the transaction. This process in known as an inner tensor product or a contraction. The vectors in question must have the same number of dimensions. The metric defines how to form a scalar as the indices are examined one−by−one. Metrics in math can be anything, but nature imposes constraints on which ones are important in physics. An aside: mathematicians require the distance is non−negative, but physicists do not. I will be using the physics notion of a metric. In looking at events in spacetime (a 4−dimensional vector), the axioms of special relativity require the Minkowski metric, which is a 4x4 real matrix which has down the diagonal 1, −1, −1, −1 and zeros elsewhere. Some people prefer the signs to be flipped, but to be consistent with everything else on this site, I choose this convention. Another popular choice is the Euclidean metric, which is the same as an identity matrix. The result of general relativity for a spherically symmetric, non−rotating mass is the Schwarzschild metric, which has "non−one" terms down the diagonal, zeros elsewhere, and becomes the Minkowski metric in the limit of the mass going to zero or the radius going to infinity.

An outer tensor product is a way to increase the rank of tensors. The tensor product of two vectors will be a 2−rank tensor. A vector can be viewed as the tensor product of a set of basis vectors.

■ What Are Quaternions?

Quaternions could be viewed as the outer tensor product of a scalar and a 3−vector. Under rotation for an event in spacetime represented by a quaternion, time is unchanged, but the 3−vector for space would be rotated. The treatment of scalars is the same as above, but the notion of vectors is far more restrictive, as restrictive as the notion of scalars. Quaternions can only handle 3−vectors. To those familiar to playing with higher dimensions, this may appear too restrictive to be of interest. Yet physics on both the quantum and cosmological scales is confined to 3−spatial dimensions. Note that the infinite Hilbert spaces in quantum mechanics a function of the principle quantum number n, not the spatial dimensions. An infinite collection of quaternions of the form (En, Pn) could represent a quantum state. The Hilbert space is formed using the Euclidean product (q* q').

A dual quaternion is formed by taking the conjugate, because q* q = (t^2 + X.X, 0). A tangent quaternion is created by having an operator act on a quaternion−valued function

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)\left(f(q), \vec{F}(q)\right) =$$
$$\left(\frac{\partial f}{\partial t} - \vec{\nabla}.\vec{F},\ \frac{\partial \vec{F}}{\partial t} + \vec{\nabla} f + \vec{\nabla} X \vec{F}\right)$$

What would happen to these five terms if space were shrunk? The 3−vector F would get shrunk, as would the divisors in the Del operator, making functions acted on by Del get larger. The scalar terms are completely unaffected by shrinking space, because df/dt has nothing to shrink, and the Del and F cancel each other. The time derivative of the 3−vector is a contravariant vector, because F would get smaller. The gradient of the scalar field is a covariant vector, because of the work of the Del operator in the divisor makes it larger. The curl at first glance might appear as a draw, but it is a covariant vector capacity because of the right−angle nature of the cross product. Note that if time where to shrink exactly as much as space, nothing in the tangent quaternion would change.

A quaternion equation must generate the same collection of tensors on both sides. Consider the product of two events, q and q':

$$(t, \vec{X})\left(t', \vec{X'}\right) = \left(t\,t' - \vec{X}.\vec{X'},\ t\,\vec{X'} + \vec{X}\,t' + \vec{X} \times \vec{X'}\right)$$
$$\text{scalars}: t,\ t',\ tt' - \vec{X}.\vec{X'}$$
$$\text{polar vectors}: \vec{X},\ \vec{X'},\ t\,\vec{X'} + \vec{X}\,t'$$
$$\text{axial vectors}: \vec{X} \times \vec{X'}$$

Where is the axial vector for the left hand side? It is imbedded in the multiplication operation, honest :−)

$$\left(t', \vec{X'}\right)(t, \vec{X}) = \left(t'\,t - \vec{X'}.\vec{X},\ t'\,\vec{X} + \vec{X'}\,t + \vec{X'} \times \vec{X}\right)$$
$$= \left(t\,t' - \vec{X}.\vec{X'},\ t\,\vec{X'} + \vec{X}\,t' - \vec{X} \times \vec{X'}\right)$$

The axial vector is the one that flips signs if the order is reversed.

Terms can continue to get more complicated. In a quaternion triple product, there will be terms of the form (XxX').X". This is called a pseudo−scalar, because it does not change under a rotation, but it will change signs under a reflection, due to the cross product. You can convince yourself of this by noting that the cross product involves the sine of an angle and the dot product involves the cosine of an angle. Neither of these will change under a rotation, and an even function times an odd function is odd. If the order of quaternion triple product is changed, this scalar will change signs for at each step in the permutation.

It has been my experience that any tensor in physics can be expressed using quaternions. Sometimes it takes a bit of effort, but it can be done.

Individual parts can be isolated if one chooses. Combinations of conjugation operators which flip the sign of a vector, and symmetric and antisymmetric products can isolate any particular term. Here are all the terms of the example from above

$$(t, \vec{X}) \left(t', \vec{X'}\right) = \left(t\,t' - \vec{X}.\vec{X'},\ t\,\vec{X'} + \vec{X}\,t' + \vec{X} \times \vec{X'}\right)$$

$$\text{scalars}: t = \frac{q + q^*}{2}, \quad t' = \frac{q' + q'^*}{2},$$

$$t t' - \vec{X}.\vec{X'} = \frac{qq' + (qq')^*}{2}$$

$$\text{polar vectors}: \vec{X} = \frac{q - q^*}{2}, \ \vec{X'} = \frac{q' - q'^*}{2},$$

$$t\,\vec{X'} + \vec{X}\,t' = \frac{(qq' + (q'q)) - (qq' + (q'q))^*}{4}$$

$$\text{axial vectors}: \vec{X} \times \vec{X'} = \frac{qq' - (q'q)}{2}$$

The metric for quaternions is imbedded in Hamilton's rule for the field.

$$\vec{i}^{\,2} = \vec{j}^{\,2} = \vec{k}^{\,2} = \vec{i}\,\vec{j}\,\vec{k} = -1$$

This looks like a way to generate scalars from vectors, but it is more than that. It also says implicitly that i j = k, j k = i, and i, j, k must have inverses. This is an important observation, because it means that inner and outer tensor products can occur in the same operation. When two quaternions are multiplied together, a new scalar (inner tensor product) and vector (outer tensor product) are formed.

How can the metric be generalized for arbitrary transformations? The traditional approach would involve playing with Hamilton's rules for the field. I think that would be a mistake, since that rule involves the fundamental definition of a quaternion. Change the rule of what a quaternion is in one context and it will not be possible to compare it to a quaternion in another context. Instead, consider an arbitrary transformation T which takes q into q'

$$q \longrightarrow q' = T\,q$$

T is also a quaternion, in fact it is equal to q' q^−1. This is guaranteed to work locally, within neighborhoods of q and q'. There is no promise that it will work globally, that one T will work for any q. Under certain circumstances, T will work for any q. The important thing to know is that a transformation T necessarily exists because quaternions are a field. The two most important theories in physics, general relativity and the standard model, involve local transformations (but the technical definition of local transformation is different than the idea presented here because it involves groups).

This quaternion definition of a transformation creates an interesting relationship between the Minkowski and Euclidean metrics.

$$\text{Let } T = I, \text{ the identity matrix}$$

$$\frac{I\,q\,I\,q + (I\,q\,I\,q)^*}{2} = (t^2 - \vec{X}.\vec{X},\ 0)$$

$$(I\,q)^*\,I\,q = (t^2 + \vec{X}.\vec{X},\ 0)$$

In order to change from wrist watch time (the interval in spacetime) to the norm of a Hilbert space does not require any change in the transformation quaternion, only a change in the multiplication step. Therefore a transformation which generates the Schwarzschild interval of general relativity should be easily portable to a Hilbert space, and that might be the start of a quantum theory of gravity.

■ So What Is the Difference?

I think it is subtle but significant. It goes back to something I learned in a graduate level class on the foundations of calculus. To make calculus rigorous requires that it is defined over a mathematical field. Physicists do this be saying that the scalars, vectors and tensors they work with are defined over the field of real or complex numbers.

What are the numbers used by nature? There are events, which consist of the scalar time and the 3−vector of space. There is mass, which is defined by the scalar energy and the 3−vector of momentum. There is the electromagnetic potential, which has a scalar field phi and a 3−vector potential A.

To do calculus with only information contained in events requires that a scalar and a 3−vector form a field. According to a theorem by Frobenius on finite dimensional fields, the only fields that fit are isomorphic to the quaternions (isomorphic is a sophisticated notion of equality, whose subtleties are appreciated only by people with a deep understanding of mathematics). To do calculus with a mass or an electromagnetic potential has an identical requirement and an identical solution. This is the logical foundation for doing physics with quaternions.

Can physics be done without quaternions? Of course it can! Events can be defined over the field of real numbers, and then the Minkowski metric and the Lorentz group can be deployed to get every result ever confirmed by experiment. Quantum mechanics can be defined using a Hilbert space defined over the field of complex numbers and return with every result measured to date.

Doing physics with quaternions is unnecessary, unless physics runs into a compatibility issue. Constraining general relativity and quantum mechanics to work within the same topological algebraic field may be the way to unite these two separately successful areas.

Inner and Outer Products of Quaternions

A good friend of mine has wondered what is means to multiply two quaternions together (this question was a hot topic in the nineteenth century). I care more about what multiplying two quaternions together can accomplish. There are two basic ways to do this: just multiply one quaternion by another, or first take the transpose of one then multiply it with the other. Each of these products can be separated into two parts: a symmetric (inner product) and an antisymmetric (outer product) components. The symmetric component will remain unchanged by exchanging the places of the quaternions, while the antisymmetric component will change its sign. Together they add up to the product. In this section, both types of inner and outer products will be formed and then related to physics.

■ The Grassman Inner and Outer Products

There are two basic ways to multiply quaternions together. There is the direct approach.

$$(t, \vec{X})\left(t', \vec{X}'\right) = \left(t\, t' - \vec{X}.\vec{X}', t\, \vec{X}' + \vec{X}\, t' + \vec{X} \times \vec{X}'\right)$$

I call this the Grassman product (I don't know if anyone else does, but I need a label). The inner product can also be called the symmetric product, because it does not change signs if the terms are reversed.

$$\text{even}\left((t, \vec{X}), \left(t', \vec{X}'\right)\right) \equiv$$

$$\equiv \frac{(t, \vec{X})\,(t', \vec{X}') + (t', \vec{X}')\,(t, \vec{X})}{2} =$$

$$\left(t\, t' - \vec{X}.\vec{X}', t\, \vec{X}' + \vec{X}\, t'\right)$$

I have defined the anticommutator (the bold curly braces) in a non−standard way, including a factor of two so I do not have to keep remembering to write it. The first term would be the Lorentz invariant interval if the two quaternions represented the same difference between two events in spacetime (i.e. t1=t2=delta t,...). The invariant interval plays a central role in special relativity. The vector terms are a frame−dependent, symmetric product of space with time and does not appear on the stage of physics, but is still a valid measurement.

The Grassman outer product is antisymmetric and is formed with a commutator.

$$\text{odd}\left((t, \vec{X}), \left(t', \vec{X}'\right)\right) \equiv$$

$$\equiv \frac{(t, \vec{X})\,(t', \vec{X}') - (t', \vec{X}')\,(t, \vec{X})}{2} = \left(0, \vec{X} \times \vec{X}'\right)$$

This is the cross product defined for two 3−vectors. It is unchanged for quaternions.

■ The Euclidean Inner and Outer Products

Another important way to multiply a pair of quaternions involves first taking the transpose of one of the quaternions. For a real−valued matrix representation, this is equivalent to multiplication by the conjugate which involves flipping the sign of the 3−vector.

$$(t, \vec{X})^* \left(t', \vec{X'}\right) = (t, -\vec{X}) \left(t', \vec{X'}\right)$$

$$= \left(t\, t' + \vec{X}.\vec{X'},\ t\, \vec{X'} - \vec{X}\, t' - \vec{X} \times \vec{X'}\right)$$

Form the Euclidean inner product.

$$\frac{(t, \vec{X})^* (t', \vec{X'}) + (t', \vec{X'})^* (t, \vec{X})}{2} = \left(t\, t' + \vec{X}.\vec{X'},\ \vec{0}\right)$$

The first term is the Euclidean norm if the two quaternions are the same (this was the reason for using the adjective "Euclidean"). The Euclidean inner product is also the standard definition of a dot product.

Form the Euclidean outer product.

$$\frac{(t, \vec{X})^* (t', \vec{X'}) - (t', \vec{X'})^* (t, \vec{X})}{2} = \left(0,\ t\, \vec{X'} - \vec{X}\, t' - \vec{X} \times \vec{X'}\right)$$

The first term is zero. The vector terms are an antisymmetric product of space with time and the negative of the cross product.

The Euclidean product is non−associative:

$$(a\, b)^*\, c \neq (a)^*\, bc$$

The norms of Euclidean products are associative because the norms are real valued:

$$|\ (a\, b)^*\, c\ | = |\ (a)^*\, bc\ |$$

The Euclidean product of quaternions might be a way t connect to the algebra of octonions,a non−associative division algebra.

■ Implications

When multiplying vectors in physics, one normally only considers the Euclidean inner product, or dot product, and the Grassman outer product, or cross product. Yet, the Grassman inner product, because it naturally generates the invariant interval, appears to play a role in special relativity. What is interesting to speculate about is the role of the Euclidean product. The Euclidean product might be a direct connection to the algebraic structure of quantum mechanics via Hilbert spaces.

Quaternion Analysis

Complex numbers are a subfield of quaternions. My hypothesis is that complex analysis should be self−evident within the structure of quaternion analysis.

The challenge is to define the derivative in a way so that a left derivative always equals a right derivative. If quaternions would only commute... Well, the scalar part of a quaternion does commute. If, in the limit, the differential element converged to a scalar, then it would commute. This idea can be defined precisely. All that is required is that the magnitude of the 3−vector goes to zero faster than the scalar. This might initially appears as an unreasonable constraint. However, there is an important application in physics. Consider a set of quaternions that represent events in spacetime. If the magnitude of the 3−space vector is less than the time scalar, events are separated by a timelike interval. It requires a speed less than the speed of light to connect the events. This is true no matter what coordinate system is chosen.

■ Defining a Quaternion

A quaternion has 4 degrees of freedom, so it needs 4 real−valued variables to be defined:

$$q = (a_0, a_1, a_2, a_3)$$

Imagine we want to do a simple binary operation such as subtraction, without having to specify the coordinate system chosen. Subtraction will only work if the coordinate systems are the same, whether it is Cartesian, spherical or otherwise. Let e0, e1, e2, and e3 be the shared, but unspecified, basis. Now we can define the difference between two quaternion q and q' that is independent of the coordinate system used for the measurement.

$$dq = q' - q = ((a_0{}' - a_0)\, e_0,\\ (a_1{}' - a_1)\, e_1 / 3,\ (a_2{}' - a_2)\, e_2 / 3,\ (a_3{}' - a_3)\, e_3 / 3)$$

What is unusual about this definition are the factors of a third. They will be necessary in order to define a holonomic equation later in this section. Hamilton gave each element parity with the others, a very reasonable approach. I have found that it is important to give the scalar and the sum of the 3−vector parity. Without this "scale" factor on the 3−vector, change in the scalar is not given its proper weight.

If dq is squared, the scalar part of the resulting quaternion forms a metric.

$$dq^{\wedge}2 = \left(da_0{}^2\, e_0{}^2 + da_1{}^2\, \frac{e_1{}^2}{9} + da_2{}^2\, \frac{e_2{}^2}{9} + da_3{}^2\, \frac{e_3{}^2}{9},\right.\\ \left. 2\, da_0\, da_1\, e_0\, \frac{e_1}{3},\ 2\, da_0\, da_2\, e_0\, \frac{e_2}{3},\ 2\, da_0\, da_3\, e_0\, \frac{e_3}{3}\right)$$

What should the connection be between the squares of the basis vectors? The amount of intrinsic curvature should be equal, so that a transformation between two basis 3−vectors does not contain a hidden bump. Should time be treated exactly like space? The Schwarzschild metric of general relativity suggests otherwise. Let e1, e2, and e3 form an independent, dimensionless, orthogonal basis for the 3−vector such that:

$$-\frac{1}{e_1{}^2} = -\frac{1}{e_2{}^2} = -\frac{1}{e_3{}^2} = e_0{}^2$$

This unusual relationship between the basis vectors is consistent with Hamilton's choice of 1, i, j, k if $e_0{}^2 = 1$. For that case, calculate the square of dq:

$$dq^2 = \left(da_0{}^2\, e_0{}^2 - \frac{da_1{}^2}{9\, e_0{}^2} - \frac{da_2{}^2}{9\, e_0{}^2} - \frac{da_3{}^2}{9\, e_0{}^2},\right.$$
$$\left. 2\, da_0\, \frac{da_1}{3},\, 2\, da_0\, \frac{da_2}{3},\, 2\, da_0\, \frac{da_3}{3}\right)$$

The scalar part is known in physics as the Minkowski interval between two events in flat spacetime. If e0^2 does not equal one, then the metric would apply to a non−flat spacetime. A metric that has been measured experimentally is the Schwarzschild metric of general relativity. Set $e_0{}^2$ = (1 – 2 = (1 − 2 GM/c^2 R), and calculate the square of dq:

$$dq^2 = \left(da_0{}^2 \left(1 - \frac{2\, G\, M}{c^2\, R}\right) - \frac{dA.dA}{9\, (1 - \frac{2\, G\, M}{c^2\, R})},\right.$$
$$\left. 2\, da_0\, \frac{da_1}{3},\, 2\, da_0\, \frac{da_2}{3},\, 2\, da_0\, \frac{da_3}{3}\right)$$

This is the Schwarzschild metric of general relativity. Notice that the 3−vector is unchanged (this may be a defining characteristic). There are very few opportunities for freedom in basic mathematical definitions. I have chosen this unusual relationships between the squares of the basis vectors to make a result from physics easy to express. Physics guides my choices in mathematical definitions :−)

■ An Automorphic Basis for Quaternion Analysis

A quaternion has 4 degrees of freedom. To completely specify a quaternion function on the manifold $\mathcal{H}^1$, it must also have four degrees of freedom. Three other linearly−independent variables involving q can be defined using conjugates combined with rotations:

$$q^* = (a_0\, e_0,\, -a_1\, e_1 / 3,\, -a_2\, e_2 / 3,\, -a_3\, e_3 / 3)$$

$$q^{*1} = (-\, a_0\, e_0,\, a_1\, e_1 / 3,\, -a_2\, e_2 / 3,\, -a_3\, e_3 / 3) = (e_1\, q\, e_1)^*$$

$$q^{*2} \equiv (-\, a_0\, e_0,\, -a_1\, e_1 / 3,\, +a_2\, e_2 / 3,\, -a_3\, e_3 / 3) = (e_2\, q\, e_2)^*$$

The conjugate as it is usually defined (q*) flips the sign of all but the scalar. The q*1 flips the signs of all but the e1 term, and q*2 all but the e2 term. The set q, q*, q*1, q*2 form the basis for quaternion analysis on the $\mathcal{H}^1$ manifold. The conjugate of a conjugate should give back the original quaternion.

$$(q^*)^* = q,\quad (q^{*1})^{*1} = q,\quad (q^{*2})^{*2} = q$$

Something subtle but perhaps directly related to spin happens looking at how the conjugates effect products:

$$(q\, q')^* = q'^*\, q^*$$

$$(q\, q')^{*1} = -\, q'^{*1}\, q^{*1}, \quad (q\, q')^{*2} = -\, q'^{*2}\, q^{*2}$$

$$(q\, q'\, q\, q')^{*1} = q'^{*1}\, q^{*1}\, q'^{*1}\, q^{*1}$$

The conjugate applied to a product brings the result directly back to the reverse order of the elements. The first and second conjugates point things in exactly the opposite way. The property of going "half way around" is reminiscent of spin. A tighter link is explored in the section on integral and half integral spin.

■ Future Timelike Derivative

Instead of the standard approach to quaternion analysis which focuses on left versus right derivatives, I concentrate on the ratio of scalars to 3−vectors. This is natural when thinking about the structure of Minkowski spacetime, where the ratio of the change in time to the change in 3−space defines five separate regions: timelike past, timelike future, lightlike past, lightlike future, and spacelike. There are no continuous Lorentz transformations to link these regions. Each region will require a separate definition of the derivative, and they will each have distinct properties. I will start with the simplest case, and look at a series of examples in detail.

Definition: The future timelike derivative:

Consider a covariant quaternion function f with a domain of H and a range of H. A future timelike derivative to be defined, the 3−vector must approach zero faster than the positive scalar. If this is not the case, then this definition cannot be used. Implementing these requirements involves two limit processes applied sequentially to a differential quaternion D. First the limit of the three vector is taken as it goes to zero, (D − D*)/2 −> 0. Second, the limit of the scalar is taken, (D + D*)/2 −> +0 (the plus zero indicates that it must be approached with a time greater than zero, in other words, from the future). The net effect of these two limit processes is that D−>0.

$$\frac{\partial f(q, q^*, q^{*1}, q^{*2})}{\partial q} =$$

$$= \text{limit as } \left(d, \vec{0}\right) \to +0 \left(\text{limit as } (d, \vec{D}) \to \left(d, \vec{0}\right) \left(f(q + (d, \vec{D}), q^*, q^{*1}, q^{*2}) - f(q, q^*, q^{*1}, q^{*2})\right) (d, \vec{D})^{-1}\right)\Big)$$

The definition is invariant under a passive transformation of the basis.

The 4 real variables a0, a1, a2, a3 can be represented by functions using the conjugates as a basis.

$$f(q, q^*, q^{*1}, q^{*2}) = a_0 = \frac{e_0\,(q + q^*)}{2}$$

$$f = a_1 = \frac{e_1\,(q + q^{*1})}{(-2/3)} = \frac{(q + q^{*1})\,e_1}{(-2/3)}$$

$$f = a_2 = \frac{e_2\,(q + q^{*2})}{(-2/3)} = \frac{(q + q^{*2})\,e_2}{(-2/3)}$$

$$f = a_3 = \frac{e_3\ (q + q^* + q^{*1} + q^{*2})}{(2 / 3)} = \frac{(q + q^* + q^{*1} + q^{*2})\ e_3}{(2 / 3)}$$

Begin with a simple example:

$$f\,(q,\ q^*,\ q^{*1},\ q^{*2}) = a_0 = \frac{e_0\ (q + q^*)}{2}$$

$$\frac{\partial\ a_0}{\partial\ q} = \frac{\partial\ a_0}{\partial\ q^*} = \lim\Big(\lim\Big((e_0\ ((q + (d,\ \vec{D}) + q^*) - (q + q^*)))$$
$$(2\ (d,\ \vec{D}))^{-1}\Big)\Big) = \frac{e_0}{2}$$

$$\frac{\partial\ a_0}{\partial\ q^{*1}} = \frac{\partial\ a_0}{\partial\ q^{*2}} = 0$$

The definition gives the expected result.

A simple approach to a trickier example:

$$f = a_1 = \frac{e_1\ (q + q^{*1})}{(-\ 2 / 3)}$$

$$\frac{\partial\ a_1}{\partial\ q} =$$

$$\frac{\partial\ a_1}{\partial\ q^{*1}} = \lim\Big(\lim\Big((e_1\ ((q + (d,\ \vec{D}) + q^{*1}) - (q + q^{*1})))$$
$$((-2 / 3)\ (d,\ \vec{D}))^{-1}\Big)\Big) = -\frac{3\,e_1}{2}$$

$$\frac{\partial\ a_1}{\partial\ q^*} = \frac{\partial\ a_1}{\partial\ q^{*2}} = 0$$

So far, the fancy double limit process has been irrelevant for these identity functions, because the differential element has been eliminated. That changes with the following example, where the e_1 is written on the right, but the result is the same.

$$f\,(q,\ q^*,\ q^{*1},\ q^{*2}) = a_1 = \frac{(q + q^{*1})\ e_1}{(-\ 2 / 3)}$$

$$\frac{\partial\ a_1}{\partial\ q} = \frac{\partial\ a_1}{\partial\ q^{*1}} =$$

$$= \lim\Big(\lim\Big(((q + (d,\ \vec{D}) + q^{*1}) - (q + q^{*1}))$$
$$e_1\ ((-2 / 3)\ (d,\ \vec{D}))^{-1}\Big)\Big) =$$

$$= \lim\Big(\lim\Big((d,\ \vec{D})\ e_1\ ((-2 / 3)\ (d,\ \vec{D}))^{-1}\Big)\Big) =$$

$$= \lim \left(\left(d, \vec{0}\right) e_1 \left((-2/3) \left(d, \vec{0}\right) \right)^{-1} \right) = -\frac{3\, e_1}{2}$$

Because the 3−vector goes to zero faster than the scalar for the differential element, after the first limit process, the remaining differential is a scalar so it commutes with any quaternion. This is what is required to dance around the e1 and lead to the cancellation.

The initial hypothesis was that complex analysis should be a self−evident subset of quaternion analysis. So this quaternion derivative should match up with the complex case, which is:

$$z = a + b\, i,\ b = (Z - Z^*) / 2\, i$$

$$\frac{\partial b}{\partial z} = -\frac{i}{2} = -\frac{\partial b}{\partial z^*}$$

These are the same result up to a factor of three. Quaternions have three imaginary axes.

The derivative of a quaternion applies equally well to polynomials.

$$\text{let } f = q^2$$

$$\frac{\partial f}{\partial q} = \lim \left(\lim \left(\left((q + (d, \vec{D}))^2 - q^2 \right) (d, \vec{D})^{-1} \right) \right) =$$

$$= \lim \Big(\lim$$

$$\left(\left(q^2 + q\, (d, \vec{D}) + (d, \vec{D})\, q + (d, \vec{D})^2 - q^2 \right) (d, \vec{D})^{-1} \right) \Big) =$$

$$= \lim \left(\lim \left(q + (d, \vec{D})\, q\, (d, \vec{D})^{-1} + (d, \vec{D}) \right) \right) =$$

$$= \lim \left(2\, q + \left(d, \vec{0}\right) \right) = 2\, q$$

This is the expected result for this polynomial. It would be straightforward to show that all polynomials gave the expected results.

Mathematicians might be concerned by this result, because if the 3−vector D goes to −D nothing will change about the quaternion derivative. This is actually consistent with principles of special relativity. For timelike separated events, right and left depend on the inertial reference frame, so a timelike derivative should not depend on the direction of the 3−vector.

■ Analytic Functions

There are 4 types of quaternion derivatives and 4 component functions. The following table describes the 16 derivatives for this set

\	a_0	a_1	a_2	a_3
$\frac{\partial}{\partial q}$	$\frac{e_0}{2}$	$\frac{e_1}{-2/3}$	$\frac{e_2}{-2/3}$	$\frac{e_3}{2/3}$
$\frac{\partial}{\partial q^*}$	$\frac{e_0}{2}$	0	0	$\frac{e_3}{2/3}$
$\frac{\partial}{\partial q^{*1}}$	0	$\frac{e_1}{-2/3}$	0	$\frac{e_3}{2/3}$
$\frac{\partial}{\partial q^{*2}}$	0	0	$\frac{e_2}{-2/3}$	$\frac{e_3}{2/3}$

This table will be used extensively to evaluate if a function is analytic using the chain rule. Let's see if the identity function $w = q$ is analytic.

$$\text{Let } w = q = \left(a_0\, e_0,\ a_1 \frac{e_1}{3},\ a_2 \frac{e_2}{3},\ a_3 \frac{e_3}{3}\right)$$

Use the chain rule to calculate the derivative will respect to each term:

$$\frac{\partial w}{\partial a_0}\ \frac{\partial a_0}{\partial q} = e_0\ \frac{e_0}{2} = \frac{1}{2}$$

$$\frac{\partial w}{\partial a_1}\ \frac{\partial a_1}{\partial q} = \frac{e_1}{3}\ \frac{e_1}{(-2/3)} = \frac{1}{2}$$

$$\frac{\partial w}{\partial a_2}\ \frac{\partial a_2}{\partial q} = \frac{e_2}{3}\ \frac{e_2}{(-2/3)} = \frac{1}{2}$$

$$\frac{\partial w}{\partial a_3}\ \frac{\partial a_3}{\partial q} = \frac{e_3}{3}\ \frac{e_3}{(2/3)} = -\frac{1}{2}$$

Use combinations of these terms to calculate the four quaternion derivatives using the chain rule.

$$\frac{\partial w}{\partial q} = \frac{\partial w}{\partial a_0}\ \frac{\partial a_0}{\partial q} + \frac{\partial w}{\partial a_1}\ \frac{\partial a_1}{\partial q} + \frac{\partial w}{\partial a_2}\ \frac{\partial a_2}{\partial q} + \frac{\partial w}{\partial a_3}\ \frac{\partial a_3}{\partial q} =$$
$$\frac{1}{2} + \frac{1}{2} + \frac{1}{2} - \frac{1}{2} = 1$$

$$\frac{\partial w}{\partial q^*} = \frac{\partial w}{\partial a_0}\ \frac{\partial a_0}{\partial q^*} + \frac{\partial w}{\partial a_3}\ \frac{\partial a_3}{\partial q^*} = \frac{1}{2} - \frac{1}{2} = 0$$

$$\frac{\partial w}{\partial q^{*1}} = \frac{\partial w}{\partial a_1}\ \frac{\partial a_1}{\partial q^{*1}} + \frac{\partial w}{\partial a_3}\ \frac{\partial a_3}{\partial q^{*1}} = \frac{1}{2} - \frac{1}{2} = 0$$

$$\frac{\partial w}{\partial q^{*2}} = \frac{\partial w}{\partial a_2}\ \frac{\partial a_2}{\partial q^{*2}} + \frac{\partial w}{\partial a_3}\ \frac{\partial a_3}{\partial q^{*2}} = \frac{1}{2} - \frac{1}{2} = 0$$

This has the derivatives expected if w=q is analytic in q.

Another test involves the Cauchy−Riemann equations. The presence of the three basis vectors changes things slightly.

$$\text{Let } u = (a_0 e_0, 0, 0, 0), \ \vec{V} = \left(0, a_1 \frac{e_1}{3}, a_2 \frac{e_2}{3}, a_3 \frac{e_3}{3}\right)$$

$$\frac{\partial u}{\partial a_0} \frac{e_1}{3} = \frac{\partial \vec{V}}{\partial a_1} e_0,$$

$$\frac{\partial u}{\partial a_0} \frac{e_2}{3} = \frac{\partial \vec{V}}{\partial a_2} e_0, \ \frac{\partial u}{\partial a_0} \frac{e_3}{3} = \frac{\partial \vec{V}}{\partial a_3} e_0$$

This also solves a holonomic equation.

$$\text{Scalar}\left(\left(\frac{\partial u}{\partial a_0}, \frac{\partial \vec{V}}{\partial a_1}, \frac{\partial \vec{V}}{\partial a_2}, \frac{\partial \vec{V}}{\partial a_3}\right) (e_0, e_1, e_2, e_3)\right) =$$

$$e_0 e_0 + \frac{e_1}{3} e_1 + \frac{e_2}{3} e_2 + \frac{e_3}{3} e_3 = 0$$

There are no off diagonal terms to compare.

This exercise can be repeated for the other identity functions. One noticeable change is in the role that the conjugate play for the basis vectors. Consider the identity function w = q*1. To show that this is analytic in q*1 requires that one always works with basis vectors of the q*1 variety.

$$\text{Let } u = (-a_0 e_0, 0, 0, 0),$$

$$\vec{V} = \left(0, a_1 \frac{e_1}{3}, -a_2 \frac{e_2}{3}, -a_3 \frac{e_3}{3}\right)$$

$$\frac{\partial u}{\partial a_0} \left(-\frac{e_1}{3}\right) = \frac{\partial \vec{V}}{\partial a_1} e_0,$$

$$\frac{\partial u}{\partial a_0} \frac{e_2}{3} = \frac{\partial \vec{V}}{\partial a_2} e_0, \ \frac{\partial u}{\partial a_0} \frac{e_3}{3} = \frac{\partial \vec{V}}{\partial a_3} e_0$$

This also solves a first conjugate holonomic equation.

$$\text{Scalar}\left(\left(\frac{\partial u}{\partial a_0}, \frac{\partial \vec{V}}{\partial a_1}, \frac{\partial \vec{V}}{\partial a_2}, \frac{\partial \vec{V}}{\partial a_3}\right) (e_0, e_1, e_2, e_3)^{*1}\right) =$$

$$-e_0 (-e_0) + \frac{e_1}{3} e_1 - \frac{-e_2}{3} e_2 - \frac{-e_3}{3} e_3 = 0$$

Power functions can be analyzed in exactly the same way:

$$\text{Let } w = q^2 = \left(a_0^2 e_0^2 + a_1^2 \frac{e_1^2}{9} + a_2^2 \frac{e_2^2}{9} + a_3^2 \frac{e_3^2}{9},\right.$$

$$\left. 2 a_0 a_1 e_0 \frac{e_1}{3}, 2 a_0 a_2 e_0 \frac{e_2}{3}, 2 a_0 a_3 e_0 \frac{e_3}{3}\right)$$

$$u = \left(a_0^2 e_0^2 + a_1^2 \frac{e_1^2}{9} + a_2^2 \frac{e_2^2}{9} + a_3^2 \frac{e_3^2}{9}, 0, 0, 0\right)$$

$$\vec{V} = \left(0,\ 2\,a_0\,a_1\,e_0\,\frac{e_1}{3},\ 2\,a_0\,a_2\,e_0\,\frac{e_2}{3},\ 2\,a_0\,a_3\,e_0\,\frac{e_3}{3}\right)$$

$$\frac{\partial u}{\partial a_0}\,\frac{e_1}{3} = \frac{2\,a_0\,e_0{}^2\,e_1}{3} = \frac{\partial \vec{V}}{\partial a_1}\,e_0$$

$$\frac{\partial u}{\partial a_0}\,\frac{e_2}{3} = \frac{2\,a_0\,e_0{}^2\,e_2}{3} = \frac{\partial \vec{V}}{\partial a_2}\,e_0$$

$$\frac{\partial u}{\partial a_0}\,\frac{e_3}{3} = \frac{2\,a_0\,e_3{}^2}{3} = \frac{\partial \vec{V}}{\partial a_3}$$

This time there are cross terms involved.

$$\frac{\partial u}{\partial a_1}\,e_0 = \frac{2\,a_1\,e_0\,e_1{}^2}{9} = \frac{\partial \vec{V}_1}{\partial a_0}\,\frac{e_1}{3}$$

$$\frac{\partial u}{\partial a_2}\,e_0 = \frac{2\,a_2\,e_0\,e_2{}^2}{9} = \frac{\partial \vec{V}_2}{\partial a_0}\,\frac{e_2}{3}$$

$$\frac{\partial u}{\partial a_3}\,e_0 = \frac{2\,a_3\,e_0\,e_3{}^2}{9} = \frac{\partial \vec{V}_3}{\partial a_0}\,\frac{e_3}{3}$$

At first glance, one might think these are incorrect, since the signs of the derivatives are suppose to be opposite. Actually they are, but it is hidden in an accounting trick :−) For example, the derivative of u with respect to a1 has a factor of e1^2, which makes it negative. The derivative of the first component of V with respect to a0 is positive. Keeping all the information about signs in the e’s makes things look non−standard, but they are not.

Note that these are three scalar equalities. The other Cauchy−Riemann equations evaluate to a single 3−vector equation. This represents four constraints on the four degrees of freedom found in quaternions to find out if a function happens to be analytic.

This also solves a holonomic equation.

$$\text{Scalar}\left(\left(\frac{\partial u}{\partial a_0},\ \frac{\partial \vec{V}}{\partial a_1},\ \frac{\partial \vec{V}}{\partial a_2},\ \frac{\partial \vec{V}}{\partial a_3}\right)(e_0,\ e_1,\ e_2,\ e_3)\right) =$$

$$= 2\,a_0\,e_0{}^3 + \frac{2\,a_0\,e_0\,e_1}{3}\,e_1 + \frac{2\,a_0\,e_0\,e_2}{3}\,e_2 + \frac{2\,a_0\,e_0\,e_3}{3}\,e_3 = 0$$

Since power series can be analytic, this should open the door to all forms of analysis. (I have done the case for the cube of q, and it too is analytic in q).

■ 4 Other Derivatives

So far, this work has only involved future timelike derivatives. There are five other regions of spacetime to cover. The simplest next case is for past timelike derivatives. The only change is in the limit, where the scalar approaches zero from below. This will make many derivatives look time symmetric, which is the case for most laws of physics.

A more complicated case involves spacelike derivatives. In the spacelike region, changes in time go to zero faster than the absolute value of the 3−vector. Therefore the order of the limit processes is reversed. This time the scalar approaches zero, then the 3−vector. This creates a problem, because after the first limit process, the differential element is (0, D), which will not commute with most quaternions. That will lead to the differential element not cancelling. The way around this is to take its norm, which is a scalar.

A spacelike differential element is defined by taking the ratio of a differential quaternion element D to its 3−vector, D − D*****. Let the norm of D approach zero. To be defined, the three vector must approach zero faster than its corresponding scalar. To make the definition non−singular everywhere, multiply by the conjugate. In the limit D D*/((D − D*)(D − D*))* approaches (1, 0), a scalar.

$$\frac{\partial f(q, q^*, q^{*1}, q^{*2})}{\partial q} \; \frac{\partial f(q, q^*, q^{*1}, q^{*2})^*}{\partial q} =$$

$$= \text{limit as } (0, \vec{D}) \to 0 \Big(\text{limit as } (d, \vec{D}) \to (0, \vec{D})$$
$$\Big((f(q + (d, \vec{D}), q^*, q^{*1}, q^{*2}) - f(q, q^*, q^{*1}, q^{*2}))$$
$$(d, \vec{D})^{-1} (f(q + (d, \vec{D}), q^*, q^{*1}, q^{*2}) -$$
$$f(q, q^*, q^{*1}, q^{*2}))^* (d, \vec{D})^{-1\,*} \Big)\Big)$$

To make this concrete, consider a simple example, f = q^2. Apply the definition:

$$\text{Norm}\left(\frac{\partial q^2}{\partial q}\right) = \text{limit } ((0, \vec{D}) \to 0 \;(\text{limit as } (d, \vec{D}) \to (0, \vec{D})$$
$$\Big(\Big(((a, \vec{B}) + (d, \vec{D}))^2 - (a, \vec{B})^2 \Big) (d, \vec{D})^{-1}$$
$$\Big(((a, \vec{B}) + (d, \vec{D}))^2 - (a, \vec{B})^2 \Big)^* (d, \vec{D})^{-1\,*} \Big)) =$$

$$= \lim \Big(((a, \vec{B}) +$$
$$(0, \vec{D})\,(a, \vec{B})\,(0, -\vec{D}) / \text{norm}((0, \vec{D})) + (0, \vec{D}))$$
$$((a, \vec{B}) + (0, \vec{D})\,(a, \vec{B})\,(0, -\vec{D}) / \text{norm}((0, \vec{D})) +$$
$$(0, \vec{D}))^* \Big) =$$

The second and fifth terms are unitary rotations of the 3−vector B. Since the differential element D could be pointed anywhere, this is an arbitrary rotation. Define:

$$(a, \vec{B}') = (0, \vec{D})\,(a, \vec{B})\,(0, -\vec{D}) / \text{norm}((0, \vec{D}))$$

Substitute, and continue:

$$= \lim \Big(((a, \vec{B}) + (a, \vec{B}') + (0, \vec{D}))$$
$$((a, \vec{B}) + (a, \vec{B}') + (0, \vec{D}))^* \Big) =$$

$$= \lim \left(4\,a\text{^}2 + 2\,\vec{B}.\vec{B} + 2\,\vec{B}.\vec{B}' + 2\,\vec{D}.\vec{B} + 2\,\vec{D}.\vec{B}',\ \vec{0}\right)$$

$$= \left(4\,a\text{^}2 + 2\,\vec{B}.\vec{B} + 2\,\vec{B}.\vec{B}',\ \vec{0}\right) \leq |\,2\,q\,|^2$$

Look at how wonderfully strange this is! The arbitrary rotation of the 3−vector B means that this derivative is bound by an inequality. If D is in direction of B, then it will be an equality, but D could also be in the opposite direction, leading to a destruction of a contribution from the 3−vector. The spacelike derivative can therefore interfere with itself. This is quite a natural thing to do in quantum mechanics. The spacelike derivative is positive definite, and could be used to define a Banach space.

Defining the lightlike derivative, where the change in time is equal to the change in space, will require more study. It may turn out that this derivative is singular everywhere, but it will require some skill to find a technically viable compromise between the spacelike and timelike derivative to synthesis the lightlike derivative.

The timelike quaternion derivative on a quaternion manifold is effectively a directional derivative along the real axis The spacelike derivative is a normed derivative The dual limit definition establishes the link between these two well−known types of derivatives.

Topological Properties of Quaternions

■ Topological Space

If we choose to work systematically through Wald's "General Relativity", the starting point is "Appendix A, Topological Spaces". Roughly, topology is the structure of relationships that do not change if a space is distorted. Some of the results of topology are required to make calculus rigorous.

In this section, I will work consistently with the set of quaternions, H^1, or just H for short. The difference between the real numbers R and H is that H is not a totally ordered set and multiplication is not commutative. These differences are not important for basic topological properties, so statements and proofs involving H are often identical to those for R.

First an open ball of quaternions needs to be defined to set the stage for an open set. Define an open ball in H of radius (r, 0) centered around a point (y, Y) [note: small letters are scalars, capital letters are 3−vectors] consisting of points (x, X) such that

$$\sqrt{((x-y,\ X-Y)^{*}\ (x-y,\ X-Y))} < (r,\ 0)$$

An open set in H is any set which can be expressed as a union of open balls.

[p. 423 translated] A quaternion topological space (H,T) consists of the set H together with a collection T of subsets of H with these properties:

1. The union of an arbitrary collection of subsets, each in T, is in T

2. The intersection of a finite number of subsets of T is in T

3. The entire set H and the empty set are in T

T is the topology on H. The subsets of H in T are open sets. Quaternions form a topology because they are what mathematicians call a metric space, since q* q evaluates to a real positive number or equals zero only if q is zero. Note: this is not the meaning of metric used by physicists. For example, the Minkowski metric can be negative or zero even if a point is not zero. To keep the same word with two meanings distinct, I will refer to one as the topological metric, the other as an interval metric. These descriptive labels are not used in general since context usually determines which one is in play.

An important component to standard approaches to general relativity is product spaces. This is how a topology for R^n is created. Events in spacetime require R^4, one place for time, three for space. Mathematicians get to make choices: what would change if work was done in R^2, R^3, or R^5? The precision of this notion, together with the freedom to make choices, makes exploring these decisions fun (for those few who can understand what is going on :−)

By working with H, product spaces are unnecessary. Events in spacetime can be members of an open set in H. Time is the scalar, space the 3−vector. There is no choice to be made.

■ Open Sets

The edges of sets will be examined by defining boundaries, open and closed sets, and the interior and closure of a set.

I am a practical guy who likes pragmatic definitions. Let the real numbers L and U represent arbitrary lower and upper bounds respectively such that L < U. For the quaternion topological space (H, T), consider an arbitrary induced topology (A, t) where x and a are elements of A. Use inequalities to define:

$$\text{an open set}: (L, 0) < (x - a)^* (x - a) < (U, 0)$$

$$\text{a closed set}: (L, 0) \leq (x - a)^* (x - a) \leq (U, 0)$$

$$\text{a half open set}: (L, 0) \leq (x - a)^* (x - a) < (U, 0)$$

$$\text{or } (L, 0) < (x - a)^* (x - a) \leq (U, 0)$$

$$\text{a boundary}: (L, 0) = (x - a)^* (x - a)$$

The union of an arbitrary collection of open sets is open.

The intersection of a finite number of open sets is open.

The union of a finite number of closed sets is closed.

The intersection of an arbitrary number of closed sets is closed.

Clearly there are connections between the above definitions

$$\text{open set union boundary} \rightarrow \text{closed set}$$

This creates complementary ideas. [Wald, p.424]

The interior of A is the union of all open sets contained within A.

The interior equals A if and only if A is open.

The closure of A is the intersection of all closed sets containing A.

The closure of A equals A if and only if A is closed.

Define a point set as the set where the lower bound equals the upper bound. The only open set that is a point set is the null set. The closed point set is H. A point set for the real numbers has only one element which is identical to the boundary. A point set for quaternions has an infinite number of elements, one of them identical to the boundary.

What are the implications for physics?

With quaternions, the existence an open set of events has nothing to do with the causality of that collection of events.

$$\text{an open set}: (L, 0) < (x - a)^* (x - a) < (U, 0)$$

$$\text{timelike events}: \text{scalar}((x - a)^2) > (0, 0)$$

$$\text{lightlike events}: \text{scalar}((x - a)^2) = (0, 0)$$

$$\text{spacelike events : scalar}\left((x-a)^2\right) < (0, 0)$$

A proper time can have exactly the same absolute value as a pure spacelike separation, so these two will be included in the same sets, whether open, closed or on a boundary.

There is no correlation the reverse way either. Take for example a collection of lightlike events. Even though they all share exactly the same interval − namely zero − their absolute value can vary all over the map, not staying within limits.

Although independent, these two ideas can be combined synergistically. Consider an open set S of timelike intervals.

$$\begin{aligned} S = \{&x,\ a\ E\ H,\ a\ \text{fixed};\ U,\\ &L\ E\ R \mid (L, 0) < (x-a)^* (x-a) < (U, 0),\\ &\text{and scalar}\left((x-a)^2\right) > 0\} \end{aligned}$$

The set S could depict a classical world history since they are causally linked and have good topological properties. A closed set of lightlike events could be a focus of quantum electrodynamics. Topology plus causality could be the key for subdividing different regions of physics.

Hausdorff Topology

This property is used to analyze compactness, something vital for rigorously establishing differentiation and integration.

[Wald p424] The quaternion topological space (H, T) is Hausdorff because for each pair of distinct points a, b E H, a not equal to b, one can find open sets Oa, Ob E T such that a E Oa, b I Ob and the intersection of Oa and Ob is the null set.

For example, find the half−way point between a and b. Let that be the radius of an open ball around the points a and b:

$$\text{let } (r, 0) = (a-b)^* (a-b) / 4$$

$$Oa = \{a,\ x\ E\ H,\ a\ \text{is fixed},\ r\ E\ R \mid (a-x)^* (a-x) < r\}$$

$$Ob = \{b,\ x\ E\ H,\ b\ \text{is fixed},\ r\ E\ R \mid (b-x)^* (b-x) < r\}$$

Neither set quite reaches the other, so their intersection is null.

■ Compact Sets

In this section, I will begin an investigation of compact sets of quaternions. I hope to share some of my insights into this subtle but significant topic.

First we need the definition of a compact set of quaternions.

[Translation of Wald p. 424] Let A be a subset of the quaternions H. Set A could be opened, closed or neither. An open cover of A is the union of open sets {Oa} that contains A. A union of open sets is open and could have an infinite number of members. A subset of {Oa} that still covers A is called a subcover. If the subcover has a finite number of elements it is called a finite subcover. The set A subset of H is compact if every open cover of A has a finite subcover.

Let’s find an example of a compact set of quaternions. Consider a set S composed of points with a finite number of absolute values:

$$S = \{x1, x2, ..., xn \text{ E } H;\ a1, a2, ..., an \text{ E } R, n \text{ is finite} \mid (x1 * x1)\hat{}.5 = (a1, 0), (x2 * x2)\hat{}.5 = (a2, 0), ...\}$$

The set S has an infinite number of members, since for any of the equalities, specifying the absolute value still leaves three degrees of freedom (if the domain had been x E R, then S would have had a finite number of elements). The set S can be covered by an open set {O} which could have an infinite number of members. There exists a subset {C} of {O} that is finite and still covers S. The subset {C} would have one member for each absolute value.

$$C = \left\{y \text{ E } \{O\}, e \text{ E } R, e > 0 \,\middle|\, (a1 - e) < \sqrt{y^* y} < (a1 + e, 0), (a2 - e) < \sqrt{y^* y} < (a2 + e, 0), ..., \text{one } y \text{ exists for each inequality}\right\}$$

Every set of quaternions composed of a finite number of absolute values like the set S is compact.

Notice that the set S is closed because it consists of a boundary without an interior. The link between compact, closed and bound set is important, and will be examined next

A compact set is a statement about the ability to find a finite number of open sets that cover a set, given any open cover. A closed set is the interior of a set plus the boundary of that set. A set is bound if there exists a real number M such that the distance between a point and any member of the set is less than M.

For quaternions with the standard topology, in order to have a finite number of open sets that cover the set, the set must necessarily include its boundary and be bound. In other words, to be compact is to be closed and bound, to be closed and bound is to be compact.

[Wald p. 425] Theorem 1 (Heine−Borel). A closed interval of quaternions S:

$$S = \left\{x \text{ E } H, a, b \text{ E } R, a < b \,\middle|\, (a, 0) \leq \sqrt{x^* x} \leq (b, 0)\right\}$$

with the standard topology on H is compact.

Wald does not provide a proof since it appears in many books on analysis. Invariably the Heine−Borel Theorem employs the domain of the real numbers, x E R. However, nothing in that proof changes by using quaternions as the domain.

[Wald p. 425] Theorem 2. Let the topology (H, T) be Hausdorff and let the set A subset of H be compact. Then A is closed.

Theorem 3. Let the topology (H, T) be compact and let the set A subset of H be closed. Then A is compact.

Combine these theorems to create a stronger statement on the compactness of subsets of quaternions H.

Theorem 4. A subset A of quaternions is compact if and only if it is closed and bounded.

The property of compactness is easily proved to be preserved under continuous maps.

Theorem 5. Let (H, T) and (H’, T’) be topological spaces. Suppose (H, T) is compact and the function f: H −> H’ is continuous. The f[H] = {h’ E H’ | h’ = f(h)} is compact. This creates a corollary by theorem 4.

Theorem 6. A continuous function from a compact topological space into H is bound and its absolute value attains a maximum and minimum values.

[end translation of Wald]

■ $\mathbb{R}^1$ versus $\mathbb{R}^n$

It is important to note that these theorems for quaternions are build directly on top of theorems for real numbers, $\mathbb{R}^1$. Only the domain needs to be changed to H^1. Wald continues with theorems on product spaces, specifically Tychonoff's Theorem, so that the above theorems can be extended to $\mathbb{R}^n$. In particular, the product space $\mathbb{R}^4$ should have the same topology as the quaternions.

Hopefully, subtlety matters in the discussion of foundations. $\mathbb{R}^4$ does not come equipped with a rule for multiplication, so it is qualitatively different from H^1, even if topologically similar to the quaternions. A similar issue arises for $\mathbb{R}^2$ and the complex number manifold $\mathbb{C}^1$.

A Quaternion Algebra Tool Set

Here is a compilation of basic algebra for quaternions. It should look very similar to complex algebra, since it contains three sets of complex numbers, t + x i, t + y j, and t + z k. To strengthen the link, and keep things looking simpler, all quaternions have been written as a pair of a scalar t and a 3−vector V, as in (t, V). All these relations have been tested in a C library and a Java quaternion calculator.

Technical note: every tool in this set can be expressed as working with a whole quaternion q. This is to show how to work with automorphic functions on a quaternion manifold.

Parts

$$\text{scalar}(q) = (q + q*)/2 = (t, 0)$$

$$\text{vector}(q) = (q - q*)/2 = (0, V)$$

Simple Algebra

$$|q| = \sqrt{(qq*)} = \left(\sqrt{t^2 + V.V}, 0\right)$$

$$\text{norm}(q) = qq* = (t^2 + V.V, 0)$$

$$\det(q) = (qq*)^2 = \left((t^2 + V.V)^2, 0\right)$$

$$\text{sum}(q, q') = q + q' = (t + t', V + V')$$

$$\text{dif}(q, q') = q - q' = (t - t', V - V')$$

$$\text{conj}(q) = q* = (t, -V)$$

$$\text{inv}(q) = q*/(qq*) = (t, -V)/(t^2 + V.V)$$

$$\text{adj}(q) = q*(qq*) = (t, -V)\,\text{norm}(q)$$

Multiplication

The Grassman product as defined here uses the same rule Hamilton developed. The Euclidean product takes the conjugate of the first of the two elements (following a tradition from quantum mechanics).

$$\text{Grassman\_product}(q, q') = qq' = (tt' - V.V', tV' + Vt' + V \times V')$$

$$\text{Grassman\_even\_product}(q, q') = \frac{qq' + q'q}{2} = (tt' - V.V', tV' + Vt')$$

$$\text{Grassman\_odd\_product}(q, q') = \frac{qq' - q'q}{2} = (0, V \times V')$$

$$\text{Euclidean\_product}(q, q') = q * q' = (tt' + V.V', tV' - Vt' - V \times V')$$

$$\text{Euclidean\_even\_product}(q, q') = \frac{q * q' + q' q *}{2} = (tt' + V.V', 0)$$

$$\text{Euclidean\_odd\_product}(q, q') = \frac{q * q' - q' q *}{2} = (0, tV' - Vt' - V \times V')$$

■ **Trigonometry**

$$\sin(q) = (\sin(t)\cosh(|V|), \cos(t)\sinh(|V|)V/|V|)$$

$$\cos(q) = (\cos(t)\cosh(|V|), -\sin(t)\sinh(|V|)V/|V|)$$

$$\tan(q) = \sin(q)/\cos(q)$$

Note: since the unit vectors of sine and cosine are the same, these two commute so the order is irrelevant.

$$\text{asin}(q) = -V/|V|\,\text{asinh}(qV/|V|)$$

$$\text{acos}(q) = -V/|V|\,\text{acosh}(q)$$

$$\text{atan}(q) = -V/|V|\,\text{atanh}(qV/|V|)$$

$$\sinh(q) = (\sinh(t)\cos(|V|), \cosh(t)\sin(|V|)V/|V|)$$

$$\cosh(q) = (\cosh(t)\cos(|V|), \sinh(t)\sin(|V|)V/|V|)$$

$$\tanh(q) = \sinh(q)/\cosh(q)$$

$$\text{asinh}(q) = \ln(q + (q\verb|^|2 + 1)\verb|^|.5)$$

$$\text{acosh}(q) = \ln(q + /-(q\verb|^|2 - 1)\verb|^|.5)$$

$$\text{atanh}(q) = .5\ln((1 + q)/(1 - q))$$

■ Powers

```
exp(q) =
  (exp(t) cos(|V|), exp(t) sin(|V|) V/|V|)

q^q' = exp(ln(q) x q')
```

■ Logs

```
ln(q) = (0.5 ln(t^2 + V.V), atan2(|V|, t) V/|V|)

log(q) = ln(q)/ln(10)
```

■ Quaternion Exponential Multiplication

```
q q' = {q, q'} + |[q, q']| exp(pi[q, q']/2|[q, q']|)

q*q' = {q*, q'} + |[q*, q']|
  exp(pi[q*, q']/2|[q*, q']|)
```

Andrew Millard suggested the result for the Grassman product.

Classical Mechanics

Newton’s Second Law

The form of Newton’s second law for three separate cases will be generated using quaternion operators acting on position quaternions. In classical mechanics, time and space are decoupled. One way that can be achieved algebraically is by having a time operator act only on space, or by space operator only act on a scalar function. I call this the "2 zero" rule: if there are two zeros in the generator of a law in physics, the law is classical.

■ Newton’s 2nd Law for an Inertial Reference Frame in Cartesian Coordinates

Define a position quaternion.

$$R = (t, \vec{R})$$

Operate on this once with the differential operator to get the velocity quaternion.

$$V = \left(\frac{d}{dt}, \vec{0}\right) (t, \vec{R}) = \left(1, \dot{\vec{R}}\right)$$

Operate on the velocity to get the classical inertial acceleration quaternion.

$$A = \left(\frac{d}{dt}, \vec{0}\right) \left(1, \dot{\vec{R}}\right) = \left(0, \bar{\vec{R}}\right)$$

This is the standard form for acceleration in Newton’s second law in an inertial reference frame. Because the reference frame is inertial, the first term is zero.

■ Newton’s 2nd Law in Polar Coordinates for a Central Force in a Plane

Repeat this process, but this time start with polar coordinates.

$$R = (t, r\,\mathrm{Cos}[\theta], r\,\mathrm{Sin}[\theta], 0)$$

The velocity in a plane.

$$V = \left(\frac{d}{dt}, \vec{0}\right) (t, r\,\mathrm{Cos}[\theta], r\,\mathrm{Sin}[\theta], 0) =$$

$$= (1, \dot{r}\,\mathrm{Cos}[\theta] - r\,\mathrm{Sin}[\theta]\,\dot{\theta}, \dot{r}\,\mathrm{Sin}[\theta] + r\,\mathrm{Cos}[\theta]\,\dot{\theta}, 0)$$

Acceleration in a plane.

$$A = \left(\frac{d}{dt}, \vec{0}\right)$$
$$(1, \dot{r}\,\mathrm{Cos}[\theta] - r\,\mathrm{Sin}[\theta]\,\dot{\theta}, \dot{r}\,\mathrm{Sin}[\theta] + r\,\mathrm{Cos}[\theta]\,\dot{\theta}, 0) =$$

$$= \left(0,\ -2\,\dot{r}\,\mathrm{Sin}[\theta]\,\dot{\theta} - r\,\mathrm{Cos}[\theta]\,\dot{\theta}^2 + \ddot{r}\,\mathrm{Cos}[\theta] - r\,\mathrm{Sin}[\theta]\,\ddot{\theta},\right.$$
$$\left. 2\,\dot{r}\,\mathrm{Cos}[\theta]\,\dot{\theta} - r\,\mathrm{Sin}[\theta]\,\dot{\theta}^2 + \ddot{r}\,\mathrm{Sin}[\theta] + r\,\mathrm{Cos}[\theta]\,\ddot{\theta},\ 0\right)$$

Not a pretty sight. For a central force, $\dot{\theta} = L/\mathrm{mr}^2$, and $\ddot{\theta} = 0$. Make these substitution and rotate the quaternion to get rid of the theta dependence.

$$A = (\mathrm{Cos}[\theta],\ 0,\ 0,\ -\mathrm{Sin}[\theta])$$
$$\left(\frac{d}{dt},\ \vec{0}\right)^2 (t,\ r\,\mathrm{Cos}[\theta],\ R\,\mathrm{Sin}[\theta],\ 0) =$$
$$= \left(0,\ \frac{L^2}{m^2\, r^3} + \ddot{r},\ \frac{2\,L\,\dot{r}}{m\, r^2},\ 0\right)$$

The second term is the acceleration in the radial direction, the third is acceleration in the theta direction for a central force in polar coordinates.

■ Newton’s 2nd Law in a Noninertial, Rotating Frame

Consider a noninertial example, with the frame rotating at an angular speed omega. The differential time operator is put into the first term of the quaternion, and the three directions for the angular speed are put in the next terms. This quaternion is then multiplied by the position quaternion to get the velocity in a rotating reference frame. Unlike the previous examples where the time t did not interfere with the calculations, here the time t must be set explicitly to zero (I wonder what that means?).

$$V = \left(\frac{d}{dt},\ \vec{\omega}\right)\ (0,\ \vec{R}) = \left(-\,\vec{\omega}.\vec{R}\ ,\ \dot{\vec{R}} + \vec{\omega}\times\vec{R}\right)$$

Operate on the velocity quaternion with the same operator.

$$A = \left(\frac{d}{dt},\ \vec{\omega}\right)\ \left(-\,\vec{\omega}.\vec{R}\ ,\ \dot{\vec{R}} + \vec{\omega}\times\vec{R}\right) =$$
$$= \left(-\,\dot{\vec{\omega}}.\vec{R}\ ,\ \ddot{\vec{R}} + 2\,\vec{\omega}\times\dot{\vec{R}} + \dot{\vec{\omega}}\times\vec{R} - \vec{\omega}.\vec{R}\,\vec{\omega}\right)$$

The first three terms of the 3−vector are the translational, coriolis, and azimuthal alterations respectively. The last term of the 3−vector may not look like the centrifugal force, but using a vector identity it can be rewritten:

$$-\,\vec{\omega}.\vec{R}\,\vec{\omega} = -\,\vec{\omega}\times(\vec{\omega}\times\vec{R}) + \vec{\omega}^2\,\vec{R}$$

If the angular velocity an the radius are orthogonal, then

$$\vec{\omega}\times(\vec{\omega}\times\vec{R}) = \vec{\omega}^2\,\vec{R}\ \mathrm{iff}\ \vec{\omega}.\vec{R} = 0$$

The scalar term is not zero. What this implies is not yet clear, but it may be related to the fact that the frame is not inertial.

■ Implications

Three forms of Newton’s second law were generated by choosing appropriate operator quaternions acting on position quaternions. It is impressive that complicated expressions in Newtonian mechanics can be encapsulated in quaternion one−line formulas. The differential time operator was decoupled from any differential space operators. This may be viewed as an operational definition of "classical" physics.

Oscillators and Waves

A professor of mine once said that everything in physics is a simple harmonic oscillator. Therefore it is necessary to get a handle on everything.

■ The Simple Harmonic Oscillator (SHO)

The differential equation for a simple harmonic oscillator in one dimension can be express with quaternion operators.

$$\left(\frac{d}{dt}, \vec{0}\right)^2 (0, x, 0, 0) + \left(0, \frac{k}{m} x, 0, 0\right) =$$
$$\left(0, \frac{d^2 x}{dt^2} + \frac{k\, x}{m}, 0, 0\right) = 0$$

This equation can be solved directly.

$$x \to C[2] \operatorname{Cos}\left[\frac{\sqrt{k}\, t}{\sqrt{m}}\right] + C[1] \operatorname{Sin}\left[\frac{\sqrt{k}\, t}{\sqrt{m}}\right]$$

Find the velocity by taking the derivative with respect to time.

$$\dot{x} \to \frac{\sqrt{k}\, C[1] \operatorname{Cos}\left[\frac{\sqrt{k}\, t}{\sqrt{m}}\right]}{\sqrt{m}} - \frac{\sqrt{k}\, C[2] \operatorname{Sin}\left[\frac{\sqrt{k}\, t}{\sqrt{m}}\right]}{\sqrt{m}}$$

■ The Damped Simple Harmonic Oscillator

Generate the differential equation for a damped simple harmonic oscillator as done above.

$$\left(\frac{d}{dt}, \vec{0}\right)^2 (0, x, 0, 0) +$$
$$\left(\frac{d}{dt}, \vec{0}\right) (0, b\, x, 0, 0) + \left(0, \frac{k}{m} x, 0, 0\right) =$$
$$= \left(0, \frac{d^2 x}{dt^2} + \frac{b\, d\, x}{dt} + \frac{k\, x}{m}, 0, 0\right) = 0$$

Solve the equation.

$$x \to C[1]\, E^{\frac{\left(-b\, m - \sqrt{-4\, k\, m + b^2\, m^2}\right) t}{2\, m}} + C[2]\, E^{\frac{\left(-b\, m + \sqrt{-4\, k\, m + b^2\, m^2}\right) t}{2\, m}}$$

■ The Wave Equation

Consider a wave traveling along the x direction. The equation which governs its motion is given by

$$\left(\frac{d}{v\,dt},\ \frac{d}{dx},\ 0,\ 0\right)^2 (0,\ 0,\ f[t\,v + x],\ 0) =$$

$$= \left(0,\ 0,\ \left(-\frac{d^2}{dx^2} + \frac{d^2}{dt^2\,v^2}\right) f[t\,v + x],\ \frac{2\,d^2\,f[t\,v + x]}{dt\,dx\,v}\right)$$

The third term is the one dimensional wave equation. The forth term is the instantaneous power transmitted by the wave.

■ Implications

Using the appropriate combinations of quaternion operators, the classical simple harmonic oscillator and wave equation were written out and solved. The functional definition of classical physics employed here is that the time operator is decoupled from any space operator. There is no reason why a similar combination of operators cannot be used when time and space operators are not decoupled. In fact, the four Maxwell equations appear to be one nonhomogeneous quaternion wave equation, and the structure of the simple harmonic oscillator appears in the Klein−Gordon equation.

Four Tests for a Conservative Force

There are four well−known, equivalent tests to determine if a force is conservative: the curl is zero, a potential function whose gradient is the force exists, all closed path integrals are zero, and the path integral between any two points is the same no matter what the path chosen. In this section, quaternion operators perform these tests on quaternion−valued forces.

■ 1. The Curl Is Zero

To make the discussion concrete, define a force quaternion F.

$$F = (0, -k x, -k y, 0)$$

The curl is the commutator of the differential operator and the force. If this is zero, the force is conservative.

$$\left[\left(\frac{d}{dt}, \vec{\nabla}\right), \vec{F}\right] = 0$$

Let the differential operator quaternion act on the force, and test if the vector components equal zero.

$$\left(\frac{d}{dt}, \nabla\right) F = (2 k, 0, 0, 0)$$

■ 2. There Exists a Potential Function for the Force

Operate on force quaternion using integration. Take the negative of the gradient of the first component. If the field quaternion is the same, the force is conservative.

$$F = \int F (dt, dx, dy, dz) =$$

$$= \int (k x\, dx + k y\, dy, -k x\, dt + k y\, dz, -k y\, dt - k x\, dz, 0) =$$

$$= \left(\frac{k x^2}{2} + \frac{k y^2}{2}, -k t x + k y z, -k t y - k x z, 0\right) =$$

$$\left(\frac{d}{dt}, \vec{\nabla}\right) \left(\frac{k\,x^2}{2} + \frac{k\,y^2}{2}, \vec{0}\right) = (0, -k\,x, -k\,y, 0)$$

This is the same force as we started with, so the scalar inside the integral is the scalar potential of this vector field. The vector terms inside the integral arise as constants of integration. They are zero if t=z=0. What role these vector terms in the potential quaternion may play, if any, is unknown to me.

■ 3. The Line Integral of Any Closed Loop Is Zero

Use any parameterization in the line integral, making sure it comes back to go.

$$\text{path} = (0, r\,\text{Cos}(t), r\,\text{Sin}(t), 0)$$

$$\int_0^{2\pi} F\,dt = 0$$

■ 4. The Line Integral Along Different Paths Is the Same

Choose any two parameterizations from A to B, and test that they are the same. These paths are from (0, r, 0, 0) to (0, −r, 2 r, 0).

$$\text{path1} = \left(0, r\,\text{Cos}(t), 2\,r\,\text{Sin}\left(\frac{t}{2}\right), 0\right)$$

$$\int_0^{2\pi} dt = -2\,k\,r^2$$

$$\text{path2} = (0, -t\,r + r, t\,r, 0)$$

$$\int_0^{2} F\,dt = -2\,k\,r^2$$

The same!

■ Implications

The four standard tests for a conservative force can be done with operator quaternions. One new avenue opened up is for doing path integrals. It would be interesting to attempt four dimensional path integrals to see where that might lead!

Special Relativity

Rotations and Dilations Create the Lorentz Group

In 1905, Einstein proposed the principles of special relativity without a deep knowledge of the mathematical structure behind the work. He had to rely on his old math teacher Minkowski to learn the theory of transformations (I do not know the details of Einstein's education, but it could make an interesting discussion :−) Eventually, Einstein understood general transformations, embodied in the work of Riemann, well enough to formulate general relativity.

A. W. Conway and L. Silberstein proposed a different mathematical structure behind special relativity in 1911 and 1912 respectively (a copy of Silberstein's work is available at quaternions.com). Cayley had observed back in 1854 that rotations in 3D could be achieved using a pair of unit quaternions having a norm of one:

$$q' = a\,q\,b \quad \text{where } a^* a = b^* b = 1$$

If this works in 3D space, why not do the 4D transformations of special relativity? It turns out that the unit quaternions must be complex−valued, or biquaternions. Is this so bad? Let me quote P.A.M. Dirac (Proc. Royal Irish Academy A, 1945, 50, p. 261):

"Quaternions themselves occupy a unique place in mathematics in that they are the most general quantities that satisfy the division axiom−−that the product of two factors cannot vanish without either factor vanishing. Biquaternions do not satisfy this axiom, and do not have any fundamental property which distinguishes them from other hyper−complex numbers. Also, they have eight components, which is rather too many for a simple scheme for describing quantities in space−time."

Just for the record: plenty of fine work has been done with biquaternions, and I do not deny the validity of any of it. Much effort has been directed toward "other hyper−complex numbers", such as Clifford algebras. I am making a choice to focus on quaternions for reasons outlined by Dirac.

Dirac took a Mobius transformation from complex analysis and tried to develop a quaternion analog. The approach is too general, and must be restricted to graft the results to the Lorentz group. I found his approach hard to follow. I needed something simpler :−)

It was quite the wait, but De Leo finally figured out a real quaternion representation of the Lorentz group (S. De Leo, "Quaternions and special relativity," J. Math. Phys., 37(6):2955−2968, 1996). He defined an operator he called "bar" which multiplied a quaternion by two quaternions on either side. He effectively did a commutator of this bar operation, which made for boosts without any terms from the cross product. It definitely is an approach that works.

■ Rotation + Dilation

Multiplication of complex numbers can be thought of as a rotation and a dilation. Conway and Silberstein's proposals only have the rotation component albeit using a complex number. An additional dilation term might allow quaternions to do the necessary work.

C. Möller wrote a general form for a Lorentz transformation using vectors ("The Theory of Relativity", QC6 F521, 1952, eq. 25). For fixed collinear coordinate systems:

$$\vec{X'} = \vec{X} + (\gamma - 1)\,(\vec{V}.\vec{X})\,\frac{\vec{V}}{|\vec{V}|^2} - \gamma\, t\, \vec{V}$$

$$t' = \gamma\, t - \gamma\,(\vec{V}.\vec{X})$$

$$\text{where } c = 1,\ \gamma = \frac{1}{\sqrt{1 - (v/c)^2}}$$

If V is only in the i direction, then

$$\vec{X'} = (\gamma\, \vec{X} - \gamma\, t\, \vec{V})\,\hat{i} + y\,\hat{j} + z\,\hat{k}$$

$$t' = \gamma\, t - \gamma\,(\vec{V}.\vec{X})$$

The additional complication to the X' equation handles velocities in different directions than i.

This has a vector equation and a scalar equation. A quaternion equation that would generate these terms must be devoid of any terms involving cross products. The symmetric product (anti−commutator) lacks the cross product;

$$\text{even}(q,\ q') = \frac{q\,q' + q'\,q}{2} = \left(t\,t' - \vec{X}.\vec{X'},\ t\,\vec{X} + \vec{X}\,t'\right)$$

Möller's equation looks like it should involve two terms, one of the form AqA (a rotation), the other Bq (a dilation).

$$q' =$$

$$q + (\gamma - 1)\,\frac{\text{even}\left(\text{even}\left(\vec{V}^*,\ q\right),\ \vec{V}\right)}{|\vec{V}|^2} + \gamma\,\text{even}\left(\vec{V}^*,\ q^*\right) =$$

$$= q + (\gamma - 1)\,\frac{\text{even}\,((\vec{V}.\vec{X},\ -t\,V),\ (0,\ \vec{V}))}{|\vec{V}|^2} +$$

$$\gamma\,\text{even}\,((0,\ -\vec{V}),\ (t,\ -\vec{X})) =$$

$$= (t,\ \vec{X}) + (\gamma - 1)\left(t,\ (\vec{V}.\vec{X})\,\frac{\vec{V}}{|\vec{V}|^2}\right) - \gamma\,((\vec{V}.\vec{X}),\ t\,\vec{V})$$

This is the general form of the Lorentz transformation presented by Möller. Real quaternions are used in a rotation and a dilation to perform the work of the Lorentz group.

■ Implications

Is this result at all interesting? A straight rewrite of Möller's equation would have been dull. What is interesting is the equation which generates the Lorentz transformation. Notice how the Lorentz transformation depends linearly on q, but the generator depends on q and q*. That may have interesting interpretations. The generator involves only symmetric products. There has been some question in the literature about whether special relativity handles rotations correctly. This is probably one of the more confusing topics in physics, so I will just let the observation stand by itself.

Two ways exist to use quaternions to do Lorentz transformations (to be discussed in the next section). The other technique relies on the property of a division algebra. There exists a quaternion L such that:

$$q' = L\,q \text{ such that}$$

$$\text{scalar}(q', q') = \text{scalar}(q, q) = t^2 - \vec{X}.\vec{X}$$

For a boost along the i direction,

$$L = \frac{q'}{q} = \frac{((\gamma t - \gamma v x, -\gamma v t + \gamma x, y, z)(t, -x, -y, -z))}{(t^2 + x^2 + y^2 + z^2)} =$$

$$= (\gamma t^2 - 2\gamma t v x + \gamma x^2 + y^2 + z^2, \gamma v(-t^2 + x^2),$$
$$t y - x z - \gamma t (y + v z) + \gamma x (v y + z),$$
$$t z + xy + \gamma t (v y - z) + \gamma x (-y + v z)) /$$
$$(t^2 + x^2 + y^2 + z^2)$$

$$\text{if } x = y = z = 0, \text{ then } L = (\gamma, -\gamma v, 0, 0)$$

$$\text{if } t = y = z = 0, \text{ then } L = (\gamma, \gamma v, 0, 0)$$

The quaternion L depends on the velocity and can depend on location in spacetime (85% of the type of problems assigned undergraduates in special relativity use an L that does not depend on location in spacetime). Some people view that as a bug, but I see it as a modern feature found in the standard model and general relativity as the demand that all symmetry is local. The existence of two approaches may be of interest in itself.

An Alternative Algebra for Lorentz Boosts: Local Transformations

Many problems in physics are expressed efficiently as differential equations whose solutions are dictated by calculus. The foundations of calculus were shown in turn to rely on the properties of fields (the mathematical variety, not the ones in physics). According to the theorem of Frobenius, there are only three finite dimensional fields: the real numbers (1D), the complex numbers (2D), and the quaternions (4D). Special relativity stresses the importance of 4−dimensional Minkowski spaces: spacetime, energy−momentum, and the electromagnetic potential. In this section, events in spacetime will be treated as the 4−dimensional field of quaternions. It will be shown that problems involving boosts along an axis of a reference frame can be solved using local quaternion transformations as apposed to the global transformations of the Lorentz group.

■ The Tools of Special Relativity

Events are represented as 4−vectors, which can be add or subtracted, or multiplied by a scalar. To form an inner product between two vectors requires the Minkowski metric, which can be represented by the following matrix (where c = 1).

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$$\{t, x, y, z\}.g_{\mu\nu}.\{t, x, y, z\} = t^2 - x^2 - y^2 - z^2$$

The Lorentz group is defined as the set of matrices that preserves the inner product of two 4−vectors. A member of this group is for boosts along the x axis, which can be easily defined.

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}$$

$$\Lambda_x = \begin{pmatrix} \gamma[\beta] & -\beta\,\gamma[\beta] & 0 & 0 \\ -\beta\,\gamma[\beta] & \gamma[\beta] & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The boosted 4−vector is

$$\Lambda_x.\{t, x, y, z\} = \left\{\frac{t}{\sqrt{1-\beta^2}} - \frac{x\,\beta}{\sqrt{1-\beta^2}},\ \frac{x}{\sqrt{1-\beta^2}} - \frac{t\,\beta}{\sqrt{1-\beta^2}},\ y,\ z\right\}$$

To demonstrate that the interval has been preserved, calculate the inner product.

$$\Lambda_x . \{t, x, y, z\} . g_\mu{}^\gamma . \Lambda_x . \{t, x, y, z\} = t^2 - x^2 - y^2 - z^2$$

Starting from a 4−vector, this is the only way to boost a reference frame along the x axis to another 4−vector and preserve the inner product. The transformation is classified as global because it depends only on the velocity and not on t, x, y, or z.

■ Using Quaternions in Special Relativity

Events will be treated as quaternions, a skew field or division algebra that is 4 dimensional. Any tool built to manipulate quaternions will also be a quaternion. In this way, although events play a different role from operators, they are made of identical mathematical fabric.

A squared quaternion is:

$$(t, \vec{X})^2 = (t^2 - \vec{X}.\vec{X}, 2 t \vec{X})$$

The first term of squaring a quaternion is the invariant interval squared. There is implicitly, a form of the Minkowski metric that is part of the rules of quaternion multiplication. The vector portion is frame−dependent. If a set of quaternions can be found that do not alter the interval, then that set would serve the same role as the Lorentz group, acting on quaternions, not on 4−vectors. If two 4−vectors x and x' are known to have the property that their intervals are identical, then the first term of squaring q[x] and q[x'] will be identical. Because quaternions are a division ring, there must exist a quaternion L such that L q[x] = q[x'] since L = q[x'] q[x]^−1. The inverse of a quaternion is its transpose over the square of the norm (which is the first term of transpose of a quaternion times itself). Apply this approach to determine L for 4−vectors boosted along the x axis.

$$L \equiv (\gamma t - \beta \gamma x, -\beta \gamma t + \gamma x, y, z)(t, x, y, z)^{-1} =$$

$$\begin{aligned}= (&\gamma t^2 + \gamma x^2 - 2 \gamma \beta t x + (y^2 + z^2), \gamma \beta (-t^2 + x^2),\\ &t (\beta \gamma z + y (1 - \gamma)) - x (\gamma \beta y + z (1 - \gamma)),\\ &t (\gamma \beta y + z (1 - \gamma)) + x (\gamma \beta z + y (1 - \gamma)))\\ &/ (t^2 + x^2 + y^2 + z^2)\end{aligned}$$

Define the Lorentz boost quaternion L along x using this equations. L depends on the relative velocity and position, making it a local, not global, transformation. See if L q[x] = q[x'].

$$\begin{aligned}&L[t, x, y, z, \beta] (t, x, y, z) =\\ &(\gamma t - \gamma \beta x, -\gamma \beta t + \gamma x, y, z)\end{aligned}$$

This is a quaternion composed of the boosted 4−vector. At this point, it can be said that _any_ problem that can be solved using 4−vectors, the Minkowski metric and a Lorentz boost along the x axis can also be solved using the above quaternion for boosting the event quaternion. This is because both techniques transform the same set of 4 numbers to the same new set of 4 numbers using the same variable beta.

Confirm the interval is unchanged.

$$(L (t, x, y, z))^2 =$$

$$= \left(t^2 - x^2 - y^2 - z^2,\ \frac{2\,(t^2\,\beta + x^2\,\beta - t\,x\,(1+\beta^2))}{-1+\beta^2},\ \frac{2\,y\,(t - x\,\beta)}{\sqrt{1-\beta^2}},\ \frac{2\,z\,(t - x\,\beta)}{\sqrt{1-\beta^2}}\right)$$

The first term is conserved as expected. The vector portion of the square is frame dependent.

■ Using Quaternions in Practice

The boost quaternion L is too complex for simple calculations. *Mathematica* does the grunge work. A great many problems in special relativity do not involve angular momentum, which in effect sets y = z = 0. Further, it is often the case that t = 0, or x = 0, or for Doppler shift problems, x = t. In these cases, the boost quaternion L becomes a very simple.

If t = 0, then

$$L = \gamma\,(1,\ \beta,\ 0,\ 0)$$

$$q \to q' = Lq$$

$$(0,\ x,\ 0,\ 0) \to (t',\ x',\ 0,\ 0) = (-\gamma\,\beta\,x,\ \gamma\,x,\ 0,\ 0)$$

If x = 0, then

$$L = \gamma\,(1,\ -\beta,\ 0,\ 0)$$

$$q \to q' = Lq$$

$$\left(t,\ \vec{0}\right) \to (t',\ x',\ 0,\ 0) = (\gamma\,t,\ -\gamma\,\beta\,t,\ 0,\ 0)$$

If t = x, then

$$L = \gamma\,(1 - \beta,\ 0,\ 0,\ 0)$$

$$q \to q' = Lq$$

$$(t,\ x,\ 0,\ 0) \to (t',\ x',\ 0,\ 0) = \gamma\,(1-\beta)\,(t,\ x,\ 0,\ 0)$$

Note: this is for blueshifts. Redshifts have a plus instead of the minus.

Over 50 problems in a sophomore−level relativistic mechanics class at MIT (8.033) have been solved using local quaternion transformations. 90% required one of these simple forms for the boost quaternion.

■ Implications

Problems in special relativity can be solved either using 4−vectors, the Minkowski metric and the Lorentz group, or using quaternions. No experimental difference between the two methods has been presented. At this point the difference is in the mathematical foundations.

An immense amount of work has gone into the study of metrics, particular in the field of general relativity. A large effort has gone into group theory and its applications to particle physics. Yet attempts to unite these two areas of study have failed.

There is no division between events, metrics and operators when solving problems using quaternions. One must be judicious in choosing quaternions that will be relevant to a particular problem in physics and therein lies the skill. Yet this creates hope that by using quaternions, the long division between metrics (the Grassman inner product) and groups of transformations (sets of quaternions that preserve the Grassman inner product) may be bridged.

Electromagnetism

Classical Electrodynamics

Maxwell speculated that someday quaternions would be useful in the analysis of electromagnetism. Hopefully after a 130 year wait, in this section we can begin that process. The Maxwell equations have been written with complex−valued quaternions back in Maxwell's time. Peter Jack was the first person to write the Maxwell equations using only real−valued quaternions. My own efforts arose a year later, independently. The approach relies on a judicious use of commutators and anticommutators.

■ The Maxwell Equations

The Maxwell equations are formed from a combinations of commutators and anticommutators of the differential operator and the electric and magnetic fields E and B respectively (for isolated charges in a vacuum.

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{B})\right) + \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{E})\right) =$$

$$\left(-\vec{\nabla}\cdot\vec{B}\ ,\ \vec{\nabla}\times\vec{E} + \frac{\partial\vec{B}}{\partial t}\right) = \left(0, \vec{0}\right)$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{B})\right) - \text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{E})\right) =$$

$$\left(\vec{\nabla}\cdot\vec{E}\ ,\ \vec{\nabla}\times\vec{B} - \frac{\partial\vec{E}}{\partial t}\right) = 4\pi\,(\rho, \vec{J})$$

$$\text{where even}(A, B) = \frac{AB + BA}{2},\ \text{odd}(A, B) = \frac{AB - BA}{2}$$

The first quaternion equation embodies the homogeneous Maxwell equations. The scalar term says that there are no magnetic monopoles. The vector term is Faraday's law. The second quaternion equation is the source term. The scalar equation is Gauss' law. The vector term is Ampere's law, with Maxwell's correction.

■ The 4−Potential A

The electric and magnetic fields are often viewed as arising from the same 4−potential A. These can also be expressed using quaternions.

$$E = \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) = \left(0, -\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi\right)$$

$$B = \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right) = (0, \vec{\nabla}\times\vec{A})$$

The electric field E is the vector part of the anticommutator of the conjugates of the differential operator and the 4−potential. The magnetic field B involves the commutator.

These forms can be directly placed into the Maxwell equations.

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) +$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right)\right) =$$

$$= \left(-\vec{\nabla}\cdot\vec{\nabla}\times\vec{A}, \frac{\partial\vec{\nabla}\times\vec{A}}{\partial t} - \vec{\nabla}\times\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\times\vec{\nabla}\phi\right) =$$

$$\left(-\vec{\nabla}\cdot\vec{B}, \frac{\partial\vec{B}}{\partial t} + \vec{\nabla}\times\vec{E}\right) = (0, \vec{0})$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, \vec{A})\right)\right) -$$

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right)\right) =$$

$$= \left(-\vec{\nabla}\cdot\vec{\nabla}\phi - \vec{\nabla}\cdot\frac{\partial\vec{A}}{\partial t}, \vec{\nabla}\times\vec{\nabla}\times\vec{A} + \frac{\partial^2\vec{A}}{\partial t^2} + \frac{\partial\vec{\nabla}\phi}{\partial t}\right) =$$

$$\left(\vec{\nabla}\cdot\vec{E}, \vec{\nabla}\times\vec{B} - \frac{\partial\vec{E}}{\partial t}\right) = 4\pi(\rho, \vec{J})$$

The homogeneous terms are formed from the sum of both orders of the commutator and anticommutator. The source terms arise from the difference of two commutators and two anticommutators. It is almost as if the even/odd operators destructively interfere to generate the homogeneous equations, while the even/even and odd/odd operators constructively interfere to describe a source.

■ The Lorentz Force

The Lorentz force is generated similarly to the source term of the Maxwell equations, but there a small game required to get the signs correct for the 4−force.

$$\text{odd}\left(\left(\gamma, \gamma\vec{\beta}\right), (0, \vec{B})\right) - \text{even}\left(\left(-\gamma, \gamma\vec{\beta}\right), (0, \vec{E})\right) =$$

$$\left(\gamma\vec{\beta}\cdot\vec{E}, \gamma\vec{E} + \gamma\vec{\beta}\times\vec{B}\right)$$

This is the covariant form of the Lorentz force.

■ Conservation Laws

The continuity equation − conservation of charge − is formed by applying the conjugate of the differential operator to the source terms of the Maxwell equations.

$$\text{scalar}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\vec{\nabla}\cdot\vec{E}, \vec{\nabla}\times\vec{B}-\frac{\partial\vec{E}}{\partial t}\right)\right) =$$

$$\frac{\partial}{\partial t}\vec{\nabla}\cdot\vec{E} - \vec{\nabla}\cdot\frac{\partial\vec{E}}{\partial t} + \vec{\nabla}\cdot\vec{\nabla}\times\vec{B} =$$

$$= \text{scalar}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), 4\pi(\rho, \vec{J})\right) = 4\pi\left(\vec{\nabla}\cdot\vec{J} + \frac{\partial\rho}{\partial t}\right)$$

The dot product of the E field and the current density plus the rate of change of the charge density must equal zero. That means that charge is conserved.

Poynting's theorem for energy conservation is formed in a very similar way, except that the conjugate of electric field is used instead of the conjugate of the differential operator.

$$\text{scalar}\left((0, -\vec{E})\left(\vec{\nabla}\cdot\vec{E}, \vec{\nabla}\times\vec{B}-\frac{\partial\vec{E}}{\partial t}\right)\right) = \vec{E}\cdot\vec{\nabla}\times\vec{B} - \vec{E}\cdot\frac{\partial\vec{E}}{\partial t}$$

$$= \text{scalar}\left((0, -\vec{E}), 4\pi(\rho, \vec{J})\right) = 4\pi\vec{E}\cdot\vec{J}$$

Additional vector identities are required before the final form is reached.

$$\vec{E}\cdot(\vec{\nabla}\times\vec{B}) = \vec{B}\cdot(\vec{\nabla}\times\vec{E}) + \vec{\nabla}\cdot(\vec{B}\times\vec{E})$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$$

$$\vec{E}.\frac{\partial\vec{E}}{\partial t} = \frac{1}{2}\left(\frac{\partial\vec{E}}{\partial t}\right)^2$$

$$\vec{B}\cdot\frac{\partial\vec{B}}{\partial t} = \frac{1}{2}\left(\frac{\partial\vec{B}}{\partial t}\right)^2$$

Use these equations to simplify to the following.

$$4\pi(\vec{E}\cdot\vec{J}, 0) = \left(-\vec{\nabla}\cdot(\vec{E}\times\vec{B}) - \frac{1}{2}\left(\frac{\partial\vec{E}}{\partial t}\right)^2 - \frac{1}{2}\left(\frac{\partial\vec{B}}{\partial t}\right)^2, 0\right)$$

This is Poynting's equation.

■ Implications

The foundations of classical electrodynamics are the Maxwell equations, the Lorentz force, and the conservation laws. In this section, these basic elements have been written as quaternion equations, exploiting the actions of commutators and anticommutators. There is an interesting link between the E field and a differential operator for generating conservation laws. More importantly, the means to generate these equations using quaternion operators has been displayed. This approach looks independent from the usual method which relies on an antisymmetric 2−rank field tensor and a U(1) connection.

Electromagnetic Field Gauges

A gauge is a measure of distance. Gauges are often chosen to make solving a particular problem easier. A few are well known: the Coulomb gauge for classical electromagnetism, the Lorenz gauge which makes electromagnetism look like a simple harmonic oscillator, and the gauge invariant form which is used in the Maxwell equations. In all these cases, the E and B field is the same, only the way it is measured is different. In this section, these are all generated using a differential quaternion operator and a quaternion electromagnetic potential.

■ The Field Tensor F in Different Gauges

The anti−symmetric 2−rank electromagnetic field tensor F has 3 properties: its trace is zero, it is antisymmetric, and it contains all the components of the E and B fields. The field used in deriving the Maxwell equations had the same information written as a quaternion:

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) - (\phi, \vec{A})\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$\left(0, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

What makes this form gauge−invariant, so no matter what the choice of gauge (involving dphi/dt and Del.A), the resulting equation is identical? It is the work of the zero! Whatever the scalar field is in the first term of the generator gets subtracted away in the second term.

Generating the field tensor F in the Lorenz gauge starting from the gauge−invariant from involves swapping the fields in the following way:

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) + (\phi, -\vec{A})}{2}\right) -$$

$$\left(\frac{(\phi, \vec{A}) - (\phi, -\vec{A})}{2}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(\frac{\partial \phi}{\partial t} + \vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

The first term of the generator involves the scalar field only, (phi, 0), and the second term involves the 3−vector field only, (0, A).

The field tensor F in the Coulomb gauge is generated by subtracting away the divergence of A, which explains why the second and third terms involve only A, even though Del.A is zero :−)

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) + \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) - (\phi, -\vec{A})}{4}\right) + \left(\frac{(\phi, -\vec{A}) - (\phi, \vec{A})}{4}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(\frac{\partial \phi}{\partial t}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

The field tensor F in the temporal gauge is quite similar to the Coulomb gauge, but some of the signs have changed to target the dphi/dt term.

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) - \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) + (\phi, -\vec{A})}{4}\right) - \left(\frac{(\phi, -\vec{A}) + (\phi, \vec{A})}{4}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(-\vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

What is the simplest expression that all of these generator share? I call it the field tensor F in the light gauge:

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

The light gauge is one sign different from the Lorenz gauge, but its generator is a simple as it gets.

■ Implications

In the quaternion representation, the gauge is a scalar generated in such a way as to not alter the 3−vector. In a lists of gauges in graduate−level quantum field theory written by Kaku, the light gauge did not make the list of the top 6 gauges. There is a reason for this. Gauges are presented as a choice for a physicist to make. The most interesting gauges have to do with a long−running popularity contest. The relationship between gauges is guessed, not written explicitly as was done here. The term that did not make the cut stands out. Perhaps some of the technical issues in quantum field theory might be tackled in this gauge using quaternions.

The Maxwell Equations in the Light Gauge: QED?

What makes a theory non−classical? Use an operational definition: a classical approach neatly separates the scalar and vector terms of a quaternion. Recall how the electric field was defined (where {A, B} is the even or symmetric product over 2, and [A, B] is the odd, antisymmetric product over two or cross product).

$$E = \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, -\vec{A})\right)\right) = \left(0, -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t}\right)$$

$$B = \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right) = (0, \vec{\nabla} X \vec{A})$$

The scalar information is explicitly discarded from the E field quaternion. In this section, the scalar field that arises will be examined and shown to be the field which gives rise to gauge symmetry. The commutators and anticommutators of this scalar and vector field do not alter the homogeneous terms of the Maxwell equations, but may explain why light is a quantized, transverse wave.

■ The E and B Fields, and the Gauge with No Name

In the previous section, the electric field was generated differently from the magnetic field, since the scalar field was discard. This time that will not be done.

$$E = \text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, -\vec{A})\right) = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla} \cdot \vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi\right)$$

$$B = \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right) = (0, \vec{\nabla} X \vec{A})$$

What is the name of the scalar field, d phi/dt − Del.A which looks like some sort of gauge? It is not the Lorenz or Landau gauge which has a plus sign between the two. It is none of the popular gauges: Coulomb (Del.A = 0), axial (Az = 0), temporal (phi = 0), Feynman, unitary...

[special note: I am now testing the interpretation that this gauge constitutes the gravitational field. See the section on Einstein’s Vision]

The standard definition of a gauge starts with an arbitrary scalar function psi. The following substitutions do not effect the resulting equations.

$$\phi \;->\; \phi' = \phi - \frac{\partial \psi}{\partial t}$$

$$\vec{A} \;->\; \vec{A}' = \vec{A} + \vec{\nabla}\psi$$

This can be written as one quaternion transformation.

$$(\phi, \vec{A}) \longrightarrow (\phi', \vec{A}') = (\phi, \vec{A}) + \left(-\frac{\partial \psi}{\partial t}, \vec{\nabla} \psi\right)$$

The goal here is to find an arbitrary scalar and a 3−vector that does the same work as the scalar function psi. Let

$$p = -\frac{\partial \psi}{\partial t} \quad \text{and} \quad \vec{\alpha} = \vec{\nabla} \psi$$

Look at how the gauge symmetry changes by taking its derivative.

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) \left(-\frac{\partial \psi}{\partial t}, \vec{\nabla} \psi\right) =$$

$$\left(-\vec{\nabla} \cdot \vec{\nabla} \psi - \frac{\partial^2 \psi}{\partial t^2}, \ \vec{\nabla} X \vec{\nabla} \psi - \vec{\nabla} \frac{\partial \psi}{\partial t} + \vec{\nabla} \frac{\partial \psi}{\partial t}\right) =$$

$$\left(\frac{\partial p}{\partial t} - \vec{\nabla} \cdot \vec{\alpha}, \ 0\right)$$

This is the gauge with no name! Call it the "light gauge". That name was chosen because if the rate of change in the scalar potential phi is equal to the spatial change of the 3−vector potential A as should be the case for a photon, the distance is zero.

■ The Maxwell Equations in the Light Gauge

The homogeneous terms of the Maxwell equations are formed from the sum of both orders of the commutator and anticommutator.

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right)\right) +$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) =$$

$$= (-\vec{\nabla} \cdot \vec{\nabla} X \vec{A}, -\vec{\nabla} X \vec{\nabla} \phi) = (0, \vec{0})$$

The source terms arise from of two commutators and two anticommutators. In the classical case discussed in the previous section, this involved a difference. Here a sum will be used because it generates a simpler differential equation.

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right)\right) -$$

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) =$$

$$= \left(\frac{\partial^2 \phi}{\partial t^2} + \vec{\nabla} \cdot \vec{\nabla} \phi, \ -\frac{\partial^2 \vec{A}}{\partial t^2} + \vec{\nabla} X (\vec{\nabla} X \vec{A}) - \vec{\nabla} \vec{\nabla} \cdot \vec{A}\right)$$

$$= \left(\frac{\partial^2 \phi}{\partial t^2} + \vec{\nabla}^2 \phi, \ -\frac{\partial^2 \vec{A}}{\partial t^2} - \vec{\nabla}^2 \vec{A}\right) = 4\pi (\rho, \vec{J})$$

Notice how the scalar and vector parts have neatly partitioned themselves. This is a wave equation, except that a sign is flipped. Here is the equation for a longitudinal wave like sound.

$$\frac{\partial^2 \vec{w}}{\partial t^2} - \vec{\nabla}^2 \vec{w} = 0$$

The second time derivative of w must be the same as Del^2 w. This has a solution which depends on sines and cosines (for simplicity, the details of initial and boundary conditions are skipped, and the infinite sum has been made finite).

$$\vec{w} = \sum_{n=0}^{\infty} \mathrm{Cos}[n \pi t] \, \mathrm{Sin}[n \pi R]$$

$$\partial_t \, \partial_t \, \vec{w} - \partial_R \, \partial_R \, \vec{w} = 0$$

Hit w with two time derivatives, and out comes −n^2 pi^2 w. Take Del^2, and that creates the same results. Thus every value of n will satisfy the longitudinal wave equation.

Now to find the solution for the sum of the second time derivative and Del^2. One of the signs must be switched by doing some operation twice. Sounds like a job for i! With quaternions, the square of a normalized 3−vector equals (−1, 0), and it is i if y = z = 0 . The solution to Maxwell's equations in the light gauge is

$$\vec{w} = \sum_{n=0}^{\infty} \mathrm{Cos}[n \pi t] \, \mathrm{Sin}[n \pi R \vec{V}]$$

$$\text{if } \vec{V}^2 = -1, \text{ then } \partial_t \, \partial_t \, \vec{w} + \partial_R \, \partial_R \, \vec{w} = 0$$

Hit this two time derivatives yields −n^2 pi^2 w. Del^2 w has all of this and the normalized phase factor V^2 = (−1, 0). V acts like an imaginary phase factor that rotates the spatial component. The sum for any n is zero (the details of the solution depend on the initial and boundary conditions).

■ Implications

The solution to the Maxwell equations in the light gauge is a superposition of waves − each with a separate value of n − where the spatial part gets rotated by the 3D analogue of i. That is a quantized, transverse wave. That's fortunate, because light is a quantized transverse wave. The equations were generated by taking the classical Maxwell equations, and making them simpler.

The Lorentz Force

The Lorentz force acts on a moving charge. The covariant form of this law is, where W is work and P is momentum:

$$\left(\frac{dW}{d\tau}, \frac{d\vec{P}}{d\tau}\right) = \gamma\, e\, \left(\vec{\beta}.\vec{E}, \vec{E} + \vec{\beta} \mathrm{X} \vec{B}\right)$$

In the classical case for a point charge, beta is zero and the E = k e/r^2, so the Lorentz force simplifies to Coulomb's law. Rewrite this in terms of the potentials phi and A.

$$\left(\frac{dW}{d\tau}, \frac{d\vec{P}}{d\tau}\right) = \gamma\, e\, \left(\beta.\left(-\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi\right), -\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\beta}\,\mathrm{X}\,(\vec{\nabla}\mathrm{X}\vec{A})\right)$$

In this section, I will look for a quaternion equation that can generate this covariant form of the Lorentz force in the Lorenz gauge. By using potentials and operators, it may be possible to create other laws like the Lorentz force, in particular, one for gravity.

■ A Quaternion Equation for the Lorentz Force

The Lorentz force is composed of two parts. First, there is the E and B fields. Generate those just as was done for the Maxwell equations

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)(\phi, \vec{A}) = \left(\frac{\partial\phi}{\partial t} - \vec{\nabla}\cdot\vec{A}, \frac{\partial\vec{A}}{\partial t} + \vec{\nabla}\phi + \vec{\nabla}\mathrm{X}\vec{A}\right)$$

Another component is the 4−velocity

$$V = \left(\gamma, \gamma\vec{\beta}\right)$$

Multiplying these two terms together creates thirteen terms, only 5 of whom belong to the Lorentz force. That should not be surprising since a bit of algebra was needed to select only the covariant terms that appear in the Maxwell equations. After some searching, I found the combination of terms required to generate the Lorentz force.

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A})\left(\gamma, -\gamma\vec{\beta}\right) - \left(\gamma, -\gamma\vec{\beta}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)(\phi, \vec{A}) =$$

$$= \gamma\left(\vec{\beta}\cdot\left(-\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi\right), -\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\beta}\,\mathrm{X}\,(\vec{\nabla}\mathrm{X}\vec{A})\right) =$$

$$\gamma\, e\left(\vec{\beta}.\vec{E}, \vec{E} + \vec{\beta}\times\vec{B}\right)$$

This combination of differential quaternion operator, quaternion potential and quaternion 4−velocity generates the covariant form of the Lorentz operator in the Lorenz gauge, minus a factor of the charge e which operates as a scalar multiplier.

■ Implications

By writing the covariant form of the Lorentz force as an operator acting on a potential, it may be possible to create other laws like the Lorentz force. For point sources in the classical limit, these new laws must have the form of Coulomb’s law, F = k e e’/r^2. An obvious candidate is Newton’s law of gravity, F = − G m m’/r^2.

The Stress Tensor of the Electromagnetic Field

I will outline a way to generate the terms of the symmetric 2−rank stress−momentum tensor of an electromagnetic field using quaternions. This method may provide some insight into what information the stress tensor contains.

Any equation written with 4−vectors can be rewritten with quaternions. A straight translation of terms could probably be automated with a computer program. What is more interesting is when an equation is generated by the product of operators acting on quaternion fields. I have found that generator equations often yield useful insights.

A tensor is a bookkeeping device designed to keep together elements that transform in a similar way. People can choose alternative bookkeeping systems, so long as the tensor behaves the same way under transformations. Using the terms as defined in "The classical theory of fields" by Landau and Life−sized, the antisymmetric 2−rank field tensor F is used to generate the stress tensor T

$$T^{ik} = \frac{1}{4\pi}\left(-F^{iL}F^{k}{}_{L} + \frac{1}{4}\delta^{ik}F_{LM}F^{LM}\right)$$

I have a practical sense of an E field (the stuff that makes my hair stand on end) and a B field (the invisible hand directing a compass), but have little sense of the field tensor F, a particular combination of the other two. Therefore, express the stress tensor T in terms of the E and B fields only:

$$T^{ik} = \begin{pmatrix} W & Sx & Sy & Sz \\ Sx & mxx & mxy & myz \\ Sy & myx & myy & myz \\ Sz & mzx & mzy & mzz \end{pmatrix}$$

$$W = \frac{1}{8\pi}\left(\vec{E}^2 + \vec{B}^2\right)$$

$$Sa = \frac{1}{4\pi}(\vec{E} \times \vec{B})$$

$$mab = \frac{1}{4\pi}\left(-Ea\,Eb - Ba\,Bb + 0.5\,\delta_{ab}\left(\vec{E}^2 + \vec{B}^2\right)\right)$$

Together, the energy density(W), Poynting's vector (Sa) and the Maxwell stress tensor (m_ab) are all the components of the stress tensor of the electromagnetic field.

■ Generating a Symmetric 2−Tensor Using Quaternions

How should one rationally go about to find a generator equation that creates these terms instead of using the month−long hunt−and−peck technique actually used? Everything is symmetric, so use the symmetric product:

$$\text{even}(q, q') = \frac{q q' + q' q}{2} = \left(t t' - \vec{X}.\vec{X}', \; t \vec{X}' + \vec{X} t'\right)$$

The fields E and B are kept separate except for the cross product in the Poynting vector. Individual directions of a field can be selected by using a unit vector Ua:

$$\text{even}(\vec{E}, Ux) = (-Ex, 0) \;\; \text{where}\; Ux = (0, 1, 0, 0)$$

The following double sum generates all the terms of the stress tensor:

$$T^{ik} =$$

$$\sum_{a=x}^{y,z} \sum_{b=x}^{y,z} \frac{1}{4\pi} \left(\left(\frac{\text{even}(Ua, Ub)}{3} - 1 \right) \frac{((0, E)^2 + (0, B)^2)}{2} - \right.$$

$$- \text{even}(E, Ua)\, \text{even}(E, Ub) - \text{even}(B, Ua)\, \text{even}(B, Ub) -$$

$$- \text{even}(\text{odd}(E, B), Ua) - \text{even}(\text{odd}(E, B), Ub) =$$

$$= (-Ex\,Ey - Ex\,Ez - Ey\,Ez - Bx\,By - Bx\,Bz - By\,Bz + Ey\,Bz - Ez\,By + Ez\,Bx - Ex\,Bz + Ex\,By - Ey\,Bx, \; 0) / 2\pi$$

The first line generates the energy density W, and part of the +0.5 delta(a, b)(E^2 + B^2) term of the Maxwell stress tensor. The rest of that tensor is generated by the second line. The third line creates the Poynting vector. Using quaternions, the net sum of these terms ends up in the scalar.

Does the generator equation have the correct properties? Switching the order of Ua and Ub leaves T unchanged, so it is symmetric. Check the trace, when Ua = Ub

$$\text{trace}(T^{ik}) =$$

$$= \sum_{a=x}^{y,z} \frac{1}{4\pi} \left(\left(\frac{\text{even}(Ua, Ua)}{3} - 1 \right) \frac{((0, E)^2 + (0, B)^2)}{2} - \text{even}(E, Ua)^2 - \text{even}(B, Ua)^2 \right) = 0$$

The trace equals zero, as it should.

The generator is composed of three parts that have different dependencies on the unit vectors: those terms that involve Ua and Ub, those that involve Ua or Ub, and those that involve neither. These are the Maxwell stress tensor, the Poynting vector and the energy density respectively. Changing the basis vectors Ua and Ub will effect these three components differently.

■ Implications

So what does the stress tensor represent? It looks like every combination of the 3−vectors E and B that avoids quadratics (like Ex^2) and over−counting cross terms. I like what I will call the "net" stress quaternion:

$$\text{net}(T^{ik}) =$$

$$= (-Ex\,Ey - Ex\,Ez - Ey\,Ez - Bx\,By - Bx\,Bz - By\,Bz + Ey\,Bz - Ez\,By + Ez\,Bx - Ex\,Bz + Ex\,By - Ey\,Bx,\ 0) / 2\pi$$

This has the same properties as an stress tensor. Since the vector is zero, it commutes with any other quaternion (this may be a reason it is so useful). Switching x terms for y terms would flip the signs of the terms produced by the Poynting vector as required, but not the others. There are no terms of the form Ex^2, which is equivalent to the statement that the trace of the tensor is zero.

On a personal note, I never thought I would understand what a symmetric 2−rank tensor was, even though I listen in on a discussion of the topic. Yes, I could nod along with the algebra, but without any sense of F, it felt hollow. Now that I have a generator and a net quaternion expression, it looks quite elegant and straightforward to me.

Quantum Mechanics

A Complete Inner Product Space with Dirac’s Bracket Notation

A mathematical connection between the bracket notation of quantum mechanics and quaternions is detailed. It will be argued that quaternions have the properties of a complete inner−product space (a Banach space for the field of quaternions). A central issue is the definition of the square of the norm. In quantum mechanics:

$$||\varphi||^2 = \langle \varphi | \varphi \rangle$$

In this section, the following assertion will be examined (* is the conjugate, so the vector flips signs):

$$||(t, \vec{X})||^2 = (t, \vec{X})^* (t, \vec{X})\ (t, \vec{X})^* (t, \vec{X})$$

The inner−product of two quaternions is defined here as the transpose (or conjugate) of the first quaternion multiplied by the second. The inner product of a function with itself is the norm.

■ The Positive Definite Norm of a Quaternion

The square of the norm of a quaternion can only be zero if every element is zero, otherwise it must have a positive value.

$$(t, \vec{X})^* (t, \vec{X}) = \left(t^2 + \vec{X}.\vec{X}, \vec{0}\right)$$

This is the standard Euclidean norm for a real 4−dimensional vector space.

The Euclidean inner−product of two quaternions can take on any value, as is the case in quantum mechanics for <phi|theta>. The adjective "Euclidean" is used to distinguish this product from the Grassman inner−product which plays a central role in special relativity (see alternative algebra for boosts).

■ Completeness

With the topology of a Euclidean norm for a real 4−dimensional vector space, quaternions are complete.

Quaternions are complete in a manner required to form a Banach space if there exists a neighborhood of any quaternion x such that there is a set of quaternions y

$$||x - y||^2 < \epsilon^4$$

for some fixed value of epsilon.

Construct such a neighborhood.

$$\left((t, \vec{X}) - \frac{\epsilon}{4}(t, \vec{X})\right)^* \left((t, \vec{X}) - \frac{\epsilon}{4}(t, \vec{X})\right)$$

$$\left((t, \vec{X}) - \frac{\epsilon}{4}(t, \vec{X})\right)^* \left((t, \vec{X}) - \frac{\epsilon}{4}(t, \vec{X})\right) =$$

$$= \left(\frac{\epsilon^4}{16}, 0, 0, 0\right) < (\epsilon^4, 0, 0, 0)$$

An infinite number of quaternions exist in the neighborhood.

Any polynomial equation with quaternion coefficients has a quaternion solution in x (a proof done by Eilenberg and Niven in 1944, cited in Birkhoff and Mac Lane's "A Survey of Modern Algebra.")

■ Identities and Inequalities

The following identities and inequalities emanate from the properties of a Euclidean norm. They are worked out for quaternions here in detail to solidify the connection between the machinery of quantum mechanics and quaternions.

The conjugate of the square of the norm equals the square of the norm of the two terms reversed.

$$< \phi \mid \varphi >^* = < \varphi \mid \phi >$$

For quaternions,

$$\left((t, \vec{X})^* \left(t', \vec{X'}\right)\right)^* = \left(t\,t' + \vec{X}.\vec{X'}, -t\,\vec{X'} + \vec{X}\,t' + \vec{X}\,x\,\vec{X'}\right)$$

$$\left(t', \vec{X'}\right)^* (t, \vec{X}) = \left(t'\,t + \vec{X'}.\vec{X}, t'\,\vec{X} - \vec{X'}\,t - \vec{X'}\,x\,\vec{X}\right)$$

These are identical, because the terms involving the cross produce will flip signs when their order changes.

For products of squares of norms in quantum mechanics,

$$< \varphi\phi \mid \varphi\phi > = < \varphi \mid \varphi > < \phi \mid \phi >$$

This is also the case for quaternions.

$$< (t, \vec{X}) \left(t', \vec{X'}\right) \Big| (t, \vec{X}) \left(t', \vec{X'}\right) > =$$

$$= \left((t, \vec{X}) \left(t', \vec{X'}\right)\right)^* (t, \vec{X}) \left(t', \vec{X'}\right)$$

$$= \left(t', \vec{X'}\right)^* (t, \vec{X})^* (t, \vec{X}) \left(t', \vec{X'}\right)$$

$$= \left(t', \vec{X'}\right)^* (t^2 + x^2 + y^2 + z^2, 0, 0, 0) \left(t', \vec{X'}\right)$$

$$= (t^2 + x^2 + y^2 + z^2, 0, 0, 0) \left(t', \vec{X'}\right)^* \left(t', \vec{X'}\right)$$

$$= (t, \vec{X})^* (t, \vec{X}) \left(t', \vec{X'}\right)^* \left(t', \vec{X'}\right)$$

$$= \; < (t, \vec{X}) \; \Big| \; (t, \vec{X}) > \; < \left(t', \vec{X'}\right) \; \Big| \; \left(t', \vec{X'}\right) >$$

The triangle inequality in quantum mechanics:

$$< \varphi + \phi \mid \phi + \varphi >^2 \; \leq \; (< \varphi \mid \varphi > + < \phi \mid \phi >)^2$$

For quaternions,

$$< (t, \vec{X}) + \left(t', \vec{X'}\right) \; \Big| \; (t, \vec{X}) + \left(t', \vec{X'}\right) >^2 =$$

$$= \left(\left(t + t', \vec{X} + \vec{X'}\right)^* \left(t + t', \vec{X} + \vec{X'}\right) \right)^2$$

$$= \left(t^2 + t'^2 + \vec{X}^2 + \vec{X'}^2 + 2\, t\, t' + 2\, \vec{X}.\vec{X'}, \; 0\right)^2$$

$$\leq$$

$$\left(t^2 + \vec{X}^2 + t'^2 + \vec{X'}^2 + \right.$$

$$\left. 2 \sqrt{(t, \vec{X})^* \; (t, \vec{X}) \; (t', \vec{X'})^* \; (t', \vec{X'})}, \; 0\right) \hat{\ } 2 \; =$$

$$\left(< (t, \vec{X}) \; \Big| \; (t, \vec{X}) > + \; < \left(t', \vec{X'}\right) \; \Big| \; \left(t', \vec{X'}\right) > \right)^2$$

If the signs of each pair of component are the same, the two sides will be equal. If the signs are different, then the cross terms will cancel on the left hand side of the inequality, making it smaller than the right hand side where terms never cancel because there are only squared terms.

The Schwarz inequality in quantum mechanics is analogous to dot products and cosines in Euclidean space.

$$| < \varphi \mid \phi > |^2 \; \leq \; < \varphi \mid \varphi > \; < \phi \mid \phi >$$

Let a third wave function, chi, be the sum of these two with an arbitrary parameter lambda.

$$\chi \; \equiv \; \varphi + \lambda\, \phi$$

The norm of chi will necessarily be greater than zero.

$$(\varphi + \lambda\, \phi)^* \; (\varphi + \lambda\, \phi) \; = \; \varphi^* \, \varphi + \lambda\, \varphi^* \, \phi \; + \; \lambda^* \, \phi^* \, \varphi \; + \; \lambda^* \, \lambda\, \phi^* \, \phi \; \geq \; 0$$

Choose the value for lambda that helps combine all the terms containing lambda.

$$\lambda \; \to \; - \frac{\phi^* \; \varphi}{\phi^* \; \phi}$$

$$\varphi^* \; \varphi \; - \; \frac{\phi^* \; \varphi\, \varphi^* \; \phi}{\phi^* \; \phi} \; \geq \; 0$$

Multiply through by the denominator, separate the two resulting terms and do some minor rearranging.

$$(\varphi^* \phi)^* \varphi^* \phi \leq \varphi^* \varphi \phi^* \phi$$

This is now the Schwarz inequality.

Another inequality:

$$2 \operatorname{Re} < \varphi \mid \phi > \leq < \varphi \mid \varphi > + < \phi \mid \phi >$$

Examine the square of the norm of the difference between two quaternions which is necessarily equal to or greater than zero.

$$0 \leq < (t, \vec{X}) - (t', \vec{X'}) \mid (t, \vec{X}) - (t', \vec{X'}) >$$

$$= \left((t - t')^2 + (\vec{X} - \vec{X'}) . (\vec{X} - \vec{X'}), \vec{0}\right)$$

The cross terms can be put on the other side of inequality, changing the sign, and leaving the sum of two norms behind.

$$\left(2 \left(t\, t' + \vec{X}.\vec{X'}\right), \vec{0}\right) \leq \left(t^2 + \vec{X}^2 + t'^2 + \vec{X'}^2, \vec{0}\right)$$

$$2 \operatorname{Re} < (t, \vec{X}) \mid (t', \vec{X'}) > \leq < (t, \vec{X}) \mid$$
$$(t, \vec{X}) > + < (t', \vec{X'}) \mid (t', \vec{X'}) >$$

The inequality holds.

The parallelogram law:

$$< \varphi + \phi \mid \phi + \varphi > + < \varphi - \phi \mid \phi - \varphi > = 2 < \varphi \mid \varphi > + 2 < \phi \mid \phi >$$

Test the quaternion norm

$$< (t, \vec{X}) + (t', \vec{X'}) \mid (t, \vec{X}) + (t', \vec{X'}) > + < (t, \vec{X}) - (t', \vec{X'}) \mid$$
$$(t, \vec{X}) - (t', \vec{X'}) > =$$

$$= \left((t + t')^2 + (\vec{X} + \vec{X'}) . (\vec{X} + \vec{X'}), \vec{0}\right) +$$
$$\left((t - t')^2 + (\vec{X} - \vec{X'}) . (\vec{X} - \vec{X'}), \vec{0}\right) =$$

$$= 2 \left(t^2 + \vec{X}^2 + t'^2 + \vec{X'}^2, \vec{0}\right) =$$

$$= 2 < (t, \vec{X}) \mid (t, \vec{X}) > + 2 < (t', \vec{X'}) \mid (t', \vec{X'}) >$$

This is twice the square of the norms of the two separate components.

■ Implications

In the case for special relativity, it was noticed that by simply squaring a quaternion, the resulting first term was the Lorentz invariant interval. From that solitary observation, the power of a mathematical field was harnessed to solve a wide range of problems in special relativity.

In a similar fashion, it is hoped that because the product of a transpose of a quaternion with a quaternion has the properties of a complete inner product space, the power of the mathematical field of quaternions can be used to solve a wide range of problems in quantum mechanics. This is an important area for further research.

Note: this goal is different from the one Stephen Adler sets out in "Quaternionic Quantum Mechanics and Quantum Fields." He tries to substitute quaternions in the place of complex numbers in the standard Hilbert space formulation of quantum mechanics. The analytical properties of quaternions do not play a critical role. It is the properties of the Hilbert space over the field of quaternions that is harnessed to solve problems. It is my opinion that since the product of a transpose of a quaternion with a quaternion already has the properties of a norm in a Hilbert space, there is no need to imbed quaternions again within another Hilbert space. I like a close shave with Occam’s razor.

Multiplying Quaternions in Polar Coordinate Form

Any quaternion can be written in polar coordinate form, which involves a scalar magnitude and angle, and a 3−vector I (which in some cases can be the more familiar i).

$$q = ||q|| \exp[\theta \vec{I}] = q^* q (\mathrm{Cos}[\theta] + \vec{I}\, \mathrm{Sin}[\theta])$$

This representation can be useful due to the properties of the exponential function, cosines and sines.

The absolute value of a quaternion is the square root of the norm, which is the transpose of a quaternion multiplied by itself.

$$|q| = \sqrt{q^* q}$$

The angle is the arccosine of the ratio of the first component of a quaternion over the norm.

$$\theta = \mathrm{ArcCos}\left(\frac{q + q^*}{2|q|}\right)$$

The vector component is generated by normalizing the pure quaternion (the final three terms) to the norm of the pure quaternion.

$$I = \frac{q - q^*}{2|q - q^*|}$$

I^2 equals −1 just like i^2. Let (0, V) = (q − q*)/2.

$$I^2 = \frac{(0, V)(0, V)}{|(0, V)||(0, V)|} = \frac{(-V.V, VxV)}{(V^2, 0)} = -1$$

It should be possible to do Fourier analysis with quaternions, and to form a Dirac delta function (or distribution). That is a project for the future. Those tools are necessary for solving problems in quantum mechanics.

■ New Method for Multiplying Quaternion Exponentials

Multiplying two exponentials is at the heart of modern analysis, whether one works with Fourier transforms or Lie groups. Given a Lie algebra of a Lie group in a sufficiently small area the identity, the product of two exponentials can be defined using the Campbell−Hausdorff formula:

$$\mathrm{Exp}[X]\, \mathrm{Exp}[y] = (X + Y) + \frac{1}{2}[X, Y](X + Y)$$
$$+ \frac{1}{12}([[X, Y], Y] - [[X, Y], X])(X + Y) + ...$$

This formula is not easy to use, and is only applicable in a small area around unity. Quaternion analysis that relies on this formula would be very limited.

I have developed (perhaps for the first time) a simpler and general way to express the product of two quaternion exponentials as the sum of two components. The product of two quaternions splits into a commuting and an anti−commuting part. The rules for multiplying commuting quaternions are identical to those for complex numbers. The anticommuting part needs to be purely imaginary. The Grassman product (q q') of two quaternion exponentials and the Euclidean product (q* q') should both have these properties. Together these define the needs for the product of two quaternion exponentials.

$$\text{Let } q = \text{Exp}[X] \quad q' = \text{Exp}[Y]$$

$$q\,q' = \{q, q'\}^* + \text{Abs}[q, q']^* \,\text{Exp}\left[\frac{\pi}{2}\,\frac{[q, q']^*}{\text{Abs}[q, q']^*}\right]$$

$$\text{where } \{q, q'\}^* \equiv \frac{q\,q' + q'^*\,q^*}{2} \text{ and } [q, q']^* \equiv q\,q' - q'^*\,q^*$$

$$q^*\,q' = \text{same as above}$$
$$\text{where } \{q, q'\} = q^*\,q' + q'^*\,q \text{ and } [q, q'] = q^*\,q' - q'^*\,q$$

I call these operators "conjugators" because they involve taking the conjugate of the two elements. Andrew Millard made the suggestion for the Grassman product that unifies these approaches nicely. What is happening here is that both commuting and anticommuting parts scale themselves appropriately. By using an exponential that has pi/2 multiplied by a normalized quaternion, this always has a zero scalar, as it must to accurately represent an anticommuting part.

Commutators and the Uncertainty Principle

Commutators and the uncertainty principle are central to quantum mechanics. Using quaternions in these roles has already been established by others (Horwitz and Biedenharn, Annals of Physics, 157:432, 1984). The first proof of the uncertainty principle I saw relied solely on the properties of complex numbers, not on physics! In this section, I will repeat that analysis, showing how commutators and an uncertainty principle arise from the properties of quaternions (or their subfield the complex numbers).

■ Commutators

Any quaternion can be written in a polar form.

$$q = (s, V) = \sqrt{q^* q}\, \mathrm{Exp}\left[\frac{s}{\sqrt{q^* q}} \frac{V}{\sqrt{V^* V}}\right]$$

This is identical to Euler's formula except that the imaginary unit vector i is replaced by the normalized 3−vector. The two are equivalent if j = k = 0. Any quaternion could be the limit of the sum of an infinite number of other quaternions expressed in a polar form. I hope to show that such a quaternion mathematically behaves like the wave function of quantum mechanics, even if the notation is different.

To simplify things, use a normalized quaternion, so that q* q = 1. Collect the normalized 3−vector together with I = V/(V* V)^.5.

The angle s/(q* q)^.5 is a real number. Any real number can be viewed as the product of two other real numbers. This seemingly irrelevant observation lends much of the flexibility seen in quantum mechanics :−) Here is the rewrite of q.

$$q = \mathrm{Exp}[a\, b\, I]$$

$$\text{where} \quad q^* q = 1, \quad a\, b = \frac{s}{\sqrt{q^* q}}, \quad I = \frac{V}{\sqrt{V^* V}}$$

The unit vector "I" could also be viewed as the product of two quaternions. For classical quantum mechanics, this additional complication is unnecessary because all "i"'s commute.

A point of clarification on notation: the same letter will be used 4 distinct ways. There are operators, A hat, which act on a quaternion wave function by multiplying by a quaternion, capital A. If the operator A hat is an observable, then it generates a real number, (a, 0), which commutes with all quaternions, whatever their form. There is also a variable with respect to a component of a quaternion, a_i, that can be used to form a differential operator.

Define a linear operator A hat that multiplies q by the quaternion A.

$$\hat{A}\, q = A\, q$$

If the operator A hat is an observable, then the quaternion A is a real number, (a, 0). A real number will commute with any quaternion. This equation is functionally equivalent to an eigenvalue equation, with A hat as an eigenvector of q and (a, 0) as the eigenvalue. However, all of the components of this equation are quaternions, not separate structures such as an operator belonging to a group and a vector. This might make a subtle but significant difference for the mathematical structure of the theory, a point that will not be investigated here.

Define a linear operator B hat that multiplies q by the quaternion B. If B hat is an observable, then this operator can be defined in terms of the scalar variable a.

$$\text{Let } \hat{B} = -I \frac{d}{da}$$

$$\hat{B} q = -I \frac{d\,\text{Exp}[a\, b\, I]}{da} = b\, q$$

Operators A and B are linear.

$$(\hat{A} + \hat{B})\, q = \hat{A} q + \hat{B} q = a\, q + b\, q = (a + b)\, q$$

$$\hat{A}\,(q + q') = \hat{A} q + \hat{A} q' = a\, q + a' q'$$

Calculate the commutator [A, B], which involves the scalar a and the derivative with respect to a.

$$[\hat{A}, \hat{B}]\, q = (\hat{A}\hat{B} - \hat{B}\hat{A})\, q = -a\, I \frac{dq}{da} + I \frac{d\,a\,q}{da}$$

$$= -a\, I \frac{dq}{da} + a\, I \frac{dq}{da} + I\, q \frac{da}{da} = I\, q$$

The commutator acting on a quaternion is equivalent to multiplying that quaternion by the normalized 3−vector I.

■ The Uncertainty Principle

Use these operators to construct things that behave like averages (expectation values) and standard deviations.

The scalar a−−generated by the observable operator A hat acting on the normalized q−−can be calculated using the Euclidean product.

$$q^* (A\, q) = q^*\, a\, q = a\, q^*\, q = a$$

It is hard to shuffle quaternions or their operators around. Real scalars commute with any quaternion and are their own conjugates. Operators that generate such scalars can move around. Look at ways to express the expectation value of A.

$$q^* (\hat{A}\, q) = q^*\, a\, q = a\, q^*\, q = a^*\, q^*\, q = (\hat{A}\, q)^*\, q = a$$

Define a new operator A' based on A whose expectation value is always zero.

$$\text{Let } A' = A - q^* (A\, q)$$

$$q^* (A'\, q) = q^* (A - q^* (A\, q))\, q) = a - a = 0$$

Define the square of the operator in a way designed to link up with the standard deviation.

$$\text{Let } DA'^2 = q^* (A'^2 q) - (q^* (A' q))^2 = q^* (A'^2 q)$$

An identical set of tools can be defined for B.

In the section on bracket notation, the Schwarz inequality for quaternions was shown.

$$\frac{A'^* B' + B'^* A'}{2} \leq \left| A' \right| \left| B' \right|$$

The Schwarz inequality applies to quaternions, not quaternion operators. If the operators A' and B' are surrounded on both sides by q and q*, then they will behave like scalars.

The left−hand side of the Schwarz inequality can be rearranged to form a commutator.

$$q^* (A'^* B' + B'^* A') q =$$

$$q^* A'^* B' q + q^* B'^* A' q = q^* a'^* B' q + q^* (-I)^* \frac{d}{d\,a} A' q =$$

$$= q^* a' B' q - q^* (-I) \frac{d}{d\,a} A' q =$$

$$q^* (A' B' - B' A') q = q^* [A', B'] q$$

The right−hand side of the Schwarz inequality can be rearranged to form the square of the standard deviation operators.

$$q^* | A' | | B' | q =$$

$$q^* A^{*\prime} A' B^{*\prime} B' q = q^* A'^2 B'^2 q = q^* DA'^2 DB'^2 q$$

Plug both of these back into the Schwarz inequality, stripping the primes and the q's which appear on both sides along the way.

$$\frac{[A, B]}{2} \leq DA^2 DB^2$$

This is the uncertainty principle for complementary observable operators.

■ Connections to Standard Notation

This quaternion exercise can be mapped to the standard notation used in physics

$$\text{bra}: \; |\psi> \; \longrightarrow q$$

$$\text{ket}: \; <\psi| \longrightarrow q^*$$

$$\text{operator}: \; A \longrightarrow A$$

$$\text{imaginary}: i \longrightarrow I$$

$$\text{commutator}: \; [A, B] \longrightarrow [A, B]$$

$$\text{norm}: \; <\psi|\psi> \; \longrightarrow q^* q$$

$$\text{expectation of } A: \; <\psi|A\psi> \; \text{maps to } q^* A q$$

$$\text{Operator A is Hermitian} \longrightarrow$$
$$(0, \vec{A}) \text{ is anti-Hermitian,}$$
$$q^* ((0, \vec{A}) q) = ((0, -\vec{A}) q)^* q$$

$$\text{The square of the standard deviation}: \delta A^2 = <\psi|A^2 \psi> - <\psi|A\psi>^2 \; \longrightarrow DA^2$$

One subtlety to note is that a quaternion operator is anti−Hermitian only if the scalar is zero. This is probably the case for classical quantum mechanics, but quantum field theory may require full quaternion operators. The proof of the uncertainty principle shown here is independent of this issue. I do not yet understand the consequence of this point.

To get to the position−momentum uncertainty equation, make these specific maps

$$A \longrightarrow X$$

$$B \longrightarrow P = i\hbar \frac{d}{dx}$$

$$I = [A, B] \longrightarrow i\hbar [X, P]$$

$$\frac{[A, B]}{2} = \frac{I}{2} \leq DA^2 DB^2 \longrightarrow \frac{[X, P]}{2} = \frac{i\hbar}{2} \leq \delta X^2 \delta P^2$$

The product of the squares of the standard deviation for position and momentum in the x−direction has a lower bound equal to half the expectation value of the commutator of those operators. The proof is in the structure of quaternions.

■ Implications

There are many interpretations of the uncertainty principle. Here, the uncertainty principle is about quaternions of the form q = Exp[a b I]. With this insight, one can see by inspection that a plane wave Exp[((Et − P.X)/hbar I], or wave packets that are superpositions of plane waves, will have four uncertainty relations, one for the scalar Et and another three for the three−part scalar P.X. This perspective should be easy to generalize.

Unifying the Representations of Integral and Half−Integral Spin

I will show how to represent both integral and half−integral spin within the same quaternion algebraic field. This involves using quaternion automorphisms. First a sketch of why this might work will be provided. Second, small rotations in a plane around two axes will be used to show how the resulting vector points in an opposite way, depending on which involution is used to construct the infinitesimal rotation. Finally, a general identity will be used to look at what happens under exchange of two quaternions in a commutator.

■ Automorphism, Rotations, and Commutators

Quaternions are formed from the direct product of a scalar and a 3−vector. Rotational operators that act on each of the 3 components of the 3−vector act like integral angular momentum. I will show that a rotation operator that acts differently on two of the three components of the 3−vector acts like half−integral spin. What happens with the scalar is irrelevant to this dimensional counting. The same rotation matrix acting on the same quaternion behaves differently depending directly on what involutions are involved.

Quaternions have 4 degrees of freedom. If we want to represent quaternions with automorphisms, 4 are required: They are the identity automorphism, the conjugate anti−automorphism, the first conjugate anti−automorphism, and the second conjugate anti−automorphism:

$$I : q \to q$$

$$* : q \to q^*$$

$$*^1 : q \to q^{*1}$$

$$*^2 : q \to q^{*2}$$

where

$$q^{*1} \equiv (e_1\, q\, e_1)^*$$

$$q^{*2} \equiv (e_2\, q\, e_2)^*$$

e1, e2, e3 are basis vectors

The most important automorphism is the identity. Life is stable around small permutations of the identity:−) The conjugate flips the signs of the each component in the 3−vector. These two automorphisms, the identity and the conjugate, treat the 3−vector as a unit. The first and second conjugate flip the signs of all terms but the first and second terms, respectively. Therefore these operators act on only the two of the three components in the 3−vector. By acting on only two of three components, a commutator will behave differently. This small difference in behavior inside a commutator is what creates the ability to represent integral and half−integral spins.

■ Small Rotations

Small rotations about the origin will now be calculated. These will then be expressed in terms of the four automorphisms discussed above.

I will be following the approach used in J. J. Sakurai's book "Modern Quantum Mechanics", chapter 3, making modifications necessary to accommodate quaternions. First, consider rotations about the origin in the z axis. Define:

$$R_{e_3=0}(\theta) \equiv \left(\cos(\theta)\, e_0,\ 0,\ 0,\ \sin(\theta)\, \frac{e_3}{3}\right)$$

$$\text{if } q = \left(0,\ a_1 \frac{e_1}{3},\ a_2 \frac{e_2}{3},\ 0\right)$$

$$R_{e_3=0}(\theta)\, q = q' = \left(0,\ (a_1 \cos(\theta) - a_2 \sin(\theta))\, e_0 \frac{e_1}{3},\right.$$
$$\left.(a_2 \cos(\theta) + a_1 \sin(\theta))\, e_0 \frac{e_2}{3},\ 0\right)$$

Two technical points. First, Sakurai considered rotations around any point along the z axis. This analysis is confined to the z axis at the origin, a significant but not unreasonable constraint. Second, these rotations are written with generalized coordinates instead of the very familiar and comfortable x, y, z. This extra effort will be useful when considering how rotations are effected by curved spacetime. This machinery is also necessary to do quaternion analysis (please see that section, it's great :−)

There are similar rotations around the first and second axes at the origin;

$$R_{e_1=0}(\theta) = \left(\cos(\theta)\, e_0,\ \sin(\theta)\, \frac{e_1}{3},\ 0,\ 0\right)$$

$$R_{e_2=0}(\theta) = \left(\cos(\theta)\, e_0,\ 0,\ \sin(\theta)\, \frac{e_2}{3},\ 0\right)$$

Consider an infinitesimal rotation for these three rotation operators. To second order in theta,

$$\sin(\theta) = \theta + O(\theta^3),\ \cos(\theta) = \left(1 - \frac{\theta^2}{2}\right) + O(\theta^3)$$

$$R_{e_1=0}(\theta << 1) = \left(\left(1 - \frac{\theta^2}{2}\right) e_0,\ \theta \frac{e_1}{3},\ 0,\ 0\right) + O(\theta^3)$$

$$R_{e_2=0}(\theta << 1) = \left(\left(1 - \frac{\theta^2}{2}\right) e_0,\ 0,\ \theta \frac{e_2}{3},\ 0\right) + O(\theta^3)$$

$$R_{e_3=0}(\theta << 1) = \left(\left(1 - \frac{\theta^2}{2}\right) e_0,\ 0,\ 0,\ \theta \frac{e_3}{3}\right) + O(\theta^3)$$

Calculate the commutator of the first two infinitesimal rotation operators to second order in theta:

$$[R_{e_1=0}, R_{e_2=0}] =$$

$$\left(\left(1 - \frac{\theta^2}{2}\right) e_0, \theta \frac{e_1}{3}, 0, 0\right) \left(\left(1 - \frac{\theta^2}{2}\right) e_0, 0, \theta \frac{e_2}{3}, 0\right) -$$

$$-\left(\left(1 - \frac{\theta^2}{2}\right) e_0, 0, \theta \frac{e_2}{3}, 0\right) \left(\left(1 - \frac{\theta^2}{2}\right) e_0, \theta \frac{e_1}{3}, 0, 0\right) =$$

$$= \left((1 - \theta^2) e_0{}^2, \theta \frac{e_0\, e_1}{3}, \theta \frac{e_0\, e_2}{3}, \theta^2 \frac{e_1\, e_2}{9}\right) -$$

$$\left((1 - \theta^2) e_0{}^2, \theta \frac{e_0\, e_1}{3}, \theta \frac{e_0\, e_2}{3}, -\theta^2 \frac{e_1\, e_2}{9}\right) =$$

$$= 2 \left(0, 0, 0, \theta^2 \frac{e_1\, e_2}{9}\right) = 2 (R_{e_3=0} (\theta^2) - R (0))$$

To second order, the commutator of infinitesimal rotations of rotations about the first two axes equals twice one rotation about the third axis given the squared angle minus a zero rotation about an arbitrary axis (a fancy way to say the identity). Now I want to write this result using anti−automorphic involutions for the small rotation operators.

$$[R^*{}_{e_1=0}, R^*{}_{e_2=0}] =$$

$$\left(\left(1 - \frac{\theta^2}{2}\right) e_0, -\theta \frac{e_1}{3}, 0, 0\right) \left(\left(1 - \frac{\theta^2}{2}\right) e_0, 0, -\theta \frac{e_2}{3}, 0\right) -$$

$$-\left(\left(1 - \frac{\theta^2}{2}\right) e_0, 0, -\theta \frac{e_2}{3}, 0\right) \left(\left(1 - \frac{\theta^2}{2}\right) e_0, -\theta \frac{e_1}{3}, 0, 0\right) =$$

$$= \left((1 - \theta^2) e_0{}^2, -\theta \frac{e_0\, e_1}{3}, -\theta \frac{e_0\, e_2}{3}, \theta^2 \frac{e_1\, e_2}{9}\right) -$$

$$\left((1 - \theta^2) e_0{}^2, -\theta \frac{e_0\, e_1}{3}, -\theta \frac{e_0\, e_2}{3}, -\theta^2 \frac{e_1\, e_2}{9}\right) =$$

$$= 2 \left(0, 0, 0, \theta^2 \frac{e_1\, e_2}{9}\right) = 2 (R_{e_3=0} (\theta^2) - R (0))$$

Nothing has changed. Repeat this exercise one last time for the first conjugate:

$$[R^{*1}{}_{e_1=0}, R^{*1}{}_{e_2=0}] = \left(-\left(1 - \frac{\theta^2}{2}\right) e_0, \theta \frac{e_1}{3}, 0, 0\right)$$

$$\left(-\left(1 - \frac{\theta^2}{2}\right) e_0, 0, -\theta \frac{e_2}{3}, 0\right) -$$

$$-\left(-\left(1 - \frac{\theta^2}{2}\right) e_0, 0, -\theta \frac{e_2}{3}, 0\right)$$

$$\left(-\left(1 - \frac{\theta^2}{2}\right) e_0, \theta \frac{e_1}{3}, 0, 0\right) =$$

$$= \left((1-\theta^2)\, e_0{}^2,\ -\theta\, \frac{e_0\, e_1}{3},\ -\theta\, \frac{e_0\, e_2}{3},\ \theta^2\, \frac{e_1\, e_2}{9} \right) - \left((1-\theta^2)\, e_0{}^2,\ -\theta\, \frac{e_0\, e_1}{3},\ -\theta\, \frac{e_0\, e_2}{3},\ -\theta^2\, \frac{e_1\, e_2}{9} \right) =$$

$$= 2\left(0,\ 0,\ 0,\ \theta^2\, \frac{e_1\, e_2}{9}\right) = -2\,(R_{e_3=0}\,(\theta^2) - R\,(0))$$

This points exactly the opposite way, even for an infinitesimal angle!

This is the kernel required to form a unified representation of integral and half integral spin. Imagine adding up a series of these small rotations, say 2 pi of these. No doubt the identity and conjugates will bring you back exactly where you started. The first and second conjugates in the commutator will point in the opposite direction. To get back on course will require another 2 pi, because the minus of a minus will generate a plus.

■ Automorphic Commutator Identities

This is a very specific example. Is there a general identity behind this work? Here it is:

$$[q,\ q'] = [q^*,\ q'^*] = [q^{*1},\ q'^{*1}]^{*1} = [q^{*2},\ q'^{*2}]^{*2}$$

It is usually a good sign if a proposal gets more subtle by generalization :−) In this case, the negative sign seen on the z axis for the first conjugate commutator is due to the action of an additional first conjugate. For the first conjugate, the first term will have the correct sign after a 2 pi journey, but the scalar, third and forth terms will point the opposite way. A similar, but not identical story applies for the second conjugate.

With the identity, we can see exactly what happens if q changes places with q’ with a commutator. Notice, I stopped right at the commutator (not including any additional conjugator). In that case:

$$[q,\ q'] = -[q',\ q] = [q^*,\ q'^*] = -[q'^*,\ q^*] =$$

$$= \left(0,\ a_2\, a_3\, \frac{e_2\, e_3}{9} + a_3\, a_2\, \frac{e_3\, e_2}{9},\ a_3\, a_1\, \frac{e_3\, e_1}{9} + a_1\, a_3\, \frac{e_1\, e_3}{9},\ a_1\, a_2\, \frac{e_1\, e_2}{9} + a_2\, a_1\, \frac{e_2\, e_1}{9}\right)$$

$$[q^{*1},\ q'^{*1}] = -[q'^{*1},\ q^{*1}] =$$

$$= \left(0,\ a_2\, a_3\, \frac{e_2\, e_3}{9} + a_3\, a_2\, \frac{e_3\, e_2}{9},\ -a_3\, a_1\, \frac{e_3\, e_1}{9} - a_1\, a_3\, \frac{e_1\, e_3}{9},\ -a_1\, a_2\, \frac{e_1\, e_2}{9} - a_2\, a_1\, \frac{e_2\, e_1}{9}\right)$$

$$[q^{*2},\ q'^{*2}] = -[q'^{*2},\ q^{*2}] =$$

$$= \left(0,\ -a_2\, a_3\, \frac{e_2\, e_3}{9} - a_3\, a_2\, \frac{e_3\, e_2}{9},\ a_3\, a_1\, \frac{e_3\, e_1}{9} + a_1\, a_3\, \frac{e_1\, e_3}{9},\ -a_1\, a_2\, \frac{e_1\, e_2}{9} - a_2\, a_1\, \frac{e_2\, e_1}{9}\right)$$

Under an exchange, the identity and conjugate commutators form a distinct group from the commutators formed with the first and second conjugates. The behavior in a commutator under exchange of the identity automorphism and the anti−automorphic conjugate are identical. The first and second conjugates are similar, but not identical.

There are also corresponding identities for the anti−commutator:

$$\{q, q'\} = \{q^*, q'^*\}^* = -\{q^{*1}, q'^{*1}\}^{*1} = -\{q^{*2}, q'^{*2}\}^{*2}$$

At this point, I don’t know how to use them, but again, the identity and first conjugates appear to behave differently that the first and second conjugates.

■ Implications

Three different operators had to be blended together to perform this feat: commutators, conjugates and rotations. These involve issue of even/oddness, mirrors, and rotations. In a commutator under exchange of two quaternions, the identity and the conjugate behave in a united way, while the first and second conjugates form a similar, but not identical set. Because this is a general quaternion identity of automorphisms, this should be very widely applicable.

Deriving A Quaternion Analog to the Schrödinger Equation

The Schrödinger equation gives the kinetic energy plus the potential (a sum also known as the Hamiltonian H) of the wave function psi, which contains all the dynamical information about a system. Psi is a scalar function with complex values.

$$H\psi = -i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2m}\nabla^2\psi + V(0, X)\psi$$

For the time−independent case, energy is written at the operator −i hbar d/dt, and kinetic energy as the square of the momentum operator, i hbar Del, over 2m. Given the potential V(0, X) and suitable boundary conditions, solving this differential equation generates a wave function psi which contains all the properties of the system.

In this section, the quaternion analog to the Schrödinger equation will be derived from first principles. What is interesting are the constraint that are required for the quaternion analog. For example, there is a factor which might serve to damp runaway terms.

■ The Quaternion Wave Function

The derivation starts from a curious place :−) Write out classical angular momentum with quaternions.

$$(0, \vec{L}) = (0, \vec{R}\times\vec{P}) = \mathrm{odd}((0, \vec{R})(0, \vec{P}))$$

What makes this "classical" are the zeroes in the scalars. Make these into complete quaternions by bringing in time to go along with the space 3−vector R, and E with the 3−vector P.

$$(t, \vec{R})(E, \vec{P}) = (Et - \vec{R}.\vec{P}, E\vec{R} + \vec{P}t + \vec{R}\times\vec{P})$$

Define a dimensionless quaternion psi that is this product over h bar.

$$\psi \equiv \frac{(t, \vec{R})(E, \vec{P})}{\hbar} = (Et - \vec{R}.\vec{P}, E\vec{R} + \vec{P}t + \vec{R}\times\vec{P})/\hbar$$

The scalar part of psi is also seen in plane wave solutions of quantum mechanics. The complicated 3−vector is a new animal, but notice it is composed of all the parts seen in the scalar, just different permutations that evaluate to 3−vectors. One might argue that for completeness, all combinations of E, t, R and P should be involved in psi, as is the case here.

Any quaternion can be expressed in polar form:

$$q = \left|q\right| e^{\arccos\left(\frac{s}{|q|}\right)\frac{\vec{V}}{|\vec{V}|}}$$

Express psi in polar form. To make things simpler, assume that psi is normalized, so |psi| = 1. The 3−vector of psi is quite complicated, so define one symbol to capture it:

$$I \equiv \frac{E\vec{R} + \vec{P}t + \vec{R} \times \vec{P}}{|E\vec{R} + \vec{P}t + \vec{R} \times \vec{P}|}$$

Now rewrite psi in polar form with these simplifications:

$$\psi = e^{(Et - \vec{R}.\vec{P})\, I/\hbar}$$

This is what I call the quaternion wave function. Unlike previous work with quaternionic quantum mechanics (see S. Adler's book "Quaternionic Quantum Mechanics"), I see no need to define a vector space with right−hand operator multiplication. As was shown in the section on bracket notation, the Euclidean product of psi (psi* psi) will have all the properties required to form a Hilbert space. The advantage of keeping both operators and the wave function as quaternions is that it will make sense to form an interacting field directly using a product such as psi psi'. That will not be done here. Another advantage is that all the equations will necessarily be invertible.

■ Changes in the Quaternion Wave Function

We cannot derive the Schrödinger equation per se, since that involves Hermitian operators that acting on a complex vector space. Instead, the operators here will be anti−Hermitian quaternions acting on quaternions. Still it will look very similar, down to the last h bar :−) All that needs to be done is to study how the quaternion wave function psi changes. Make the following assumptions.

1. Energy and Momentum are conserved.

$$\frac{\partial E}{\partial t} = 0 \text{ and } \frac{\partial \vec{P}}{\partial t} = 0$$

2. Energy is evenly distributed in space

$$\vec{\nabla} E = 0$$

3. The system is isolated

$$\vec{\nabla} \times \vec{P} = 0$$

4. The position 3−vector X is in the same direction as the momentum 3−vector P

$$\frac{X.P}{|X||P|} = 1 \text{ which implies } \frac{de^{I}}{dt} = 0 \text{ and } \vec{\nabla} \times e^{I} = 0$$

The implications of this last assumption are not obvious but can be computed directly by taking the appropriate derivative. Here is a verbal explanation. If energy and momentum are conserved, they will not change in time. If the position 3−vector which does change is always in the same direction as the momentum 3−vector, then I will remain constant in time. Since I is in the direction of X, its curl will be zero.

This last constraint may initially appear too confining. Contrast this with the typical classical quantum mechanics. In that case, there is an imaginary factor i which contains no information about the system. It is a mathematical tool tossed in so that the equation has the correct properties. With quaternions, I is determined directly from E, t, P and X. It must be richer in information content. This particular constraint is a reflection of that.

Now take the time derivative of psi.

$$\frac{\partial \psi}{\partial t} = \frac{E\,I}{\hbar} \frac{\psi}{\sqrt{1 + \left(\frac{Et - \vec{R}.\vec{P}}{\hbar}\right)^2}}$$

The denominator must be at least 1, and can be greater than that. It can serve as a damper, a good thing to tame runaway terms. Unfortunately, it also makes solving explicitly for energy impossible unless Et − P.X equals zero. Since the goal is to make a direct connection to the Schrödinger equation, make one final assumption:

5. Et − R.P = 0

$$Et - \vec{R}.\vec{P} = 0$$

There are several important cases when this will be true. In a vacuum, E and P are zero. If this is used to study photons, then t = |R| and E = |P|. If this number happens to be constant in time, then this equation will apply to the wave front.

$$\text{if } \frac{\partial Et - \vec{R}.\vec{P}}{\partial t} = 0,\ E = \frac{\partial \vec{R}}{\partial t}.\vec{P} \text{ or } \frac{\partial \vec{R}}{\partial t} = \frac{E}{\vec{P}}$$

Now with these 5 assumptions in hand, energy can be defined with an operator.

$$\frac{\partial \psi}{\partial t} = \frac{E\,I}{\hbar}\,\psi$$

$$-I\hbar \frac{\partial \psi}{\partial t} = E\,\psi \text{ or } E = -I\hbar \frac{\partial}{\partial t}$$

The equivalence of the energy E and this operator is called the first quantization.

Take the spatial derivative of psi using the under the same assumptions:

$$\vec{\nabla}\psi = -\frac{\vec{P}\,I}{\hbar} \frac{\psi}{\sqrt{1 + \left(\frac{Et - \vec{R}.\vec{P}}{\hbar}\right)^2}}$$

$$\vec{I}\,\hbar\,\vec{\nabla}\psi = \vec{P}\,\psi \text{ or } \vec{P} = I\hbar\,\vec{\nabla}$$

Square this operator.

$$\vec{P}^2 = (mv)^2 = 2\,m\,\frac{mv^2}{2} = 2\,m\,KE = -\hbar^2\,\vec{\nabla}^2$$

The Hamiltonian equals the kinetic energy plus the potential energy.

$$\vec{H}\,\psi = -\vec{I}\,\hbar\,\frac{\partial \psi}{\partial t} = -\hbar^2\,\vec{\nabla}^2\,\psi + V\,\psi$$

Typographically, this looks very similar to the Schrödinger equation. Capital I is a normalized 3−vector, and a very complicated one at that if you review the assumptions that got us here. Phi is not a vector, but is a quaternion. This give the equation more, not less, analytical power. With all of the constraints in place, I expect that this equation will behave exactly like the Schrodinger equation. As the constraints are removed, this proposal becomes richer. There is a damper to quench runaway terms. The 3−vector I becomes quite the nightmare to deal with, but it should be possible, given we are dealing with a topological algebraic field.

■ Implications

Any attempt to shift the meaning of an equation as central to modern physics had first be able to regenerate all of its results. I believe that the quaternion analog to Schrödinger equation under the listed constraints will do the task. These is an immense amount of work needed to see as the constraints are relaxed, whether the quaternion differential equations will behave better. My sense at this time is that first quaternion analysis as discussed earlier must be made as mathematically solid as complex analysis. At that point, it will be worth pushing the envelope with this quaternion equation. If it stands on a foundation as robust as complex analysis, the profound problems seen in quantum field theory stand a chance of fading away into the background.

Introduction to Relativistic Quantum Mechanics

The relativistic quantum mechanic equation for a free particle is the Klein−Gordon equation (h=c=1)

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2 \right) \Psi = 0$$

The Schrödinger equation results from the non−relativistic limit of this equation. In this section, the machinery of the Klein−Gordon equation will be ported to quaternions.

■ The Wave Function

The wave function is the superposition of all possible states of a system. The product of the conjugate of a wave function with another wave function forms a complete inner product space. In the energy/momentum representation, this would involve all possible energy levels and momenta.

$$\Psi \equiv \text{the sum from } n = 0 \text{ to infinity of } (E_n, \vec{P}_n)$$

This infinite sum of quaternions should contain all the information about a system. The quaternion wave function can be normalized.

$$\sum_{n=0}^{\infty} (E_n, \vec{P}_n)^* (E_n, \vec{P}_n) = \sum_{n=0}^{\infty} \left(E_n^2 + \vec{P}_n^2, 0 \right) = 1$$

The first quaternion is the conjugate or transpose of the second. Since the transpose of a quaternion wave function times a wave function creates a Euclidean norm, this representation of wave functions as an infinite sum of quaternions can form a complete, normed product space.

■ The Klein−Gordon Equation

The Klein−Gordon equation can be divided into two operators that act on the wave function: the D’Alembertian and the scalar m^2. The quaternion operator required to create the D’Alembertian, along with vector identities, has already been worked out for the Maxwell equations in the Lorenz gauge.

$$\sum_{n=0}^{\infty} \left(\left(\frac{\partial}{\partial t}, \vec{\nabla} \right)^2 + \left(\frac{\partial}{\partial t}, -\vec{\nabla} \right)^2 \right) (E_n, \vec{P}_n) / 2 =$$

$$= \sum_{n=0}^{\infty} \left(\frac{\partial^2 E_n}{\partial t^2} - \vec{\nabla} \cdot \vec{\nabla} E_n - \vec{\nabla} \cdot \vec{\nabla} X \vec{P}_n, \right.$$

$$\left. \frac{\partial^2 \vec{P}_n}{\partial t^2} - \vec{\nabla} \vec{\nabla} \cdot \vec{P}_n + \vec{\nabla} X \vec{\nabla} X \vec{P}_n + \vec{\nabla} X \vec{\nabla} E_n \right)$$

The first term of the scalar, and the second term of the vector, are both equal to zero. What is left is the D'Alembertian operator acting on the quaternion wave function.

To generate the scalar multiplier m^2, substitute En and Pn for the operators d/dt and del respectively, and repeat. Since the structure of the operator is identical to the previous one, instead of the D'Alembertian times the wave function, there is En^2−Pn^2. The sum of all these terms becomes m^2.

Set the sum of these two operators equal to zero to form the Klein−Gordon equation.

$$\sum_{n=0}^{\infty} \left(\left(\frac{\partial}{\partial t}, \vec{\nabla} \right)^2 + \left(\frac{\partial}{\partial t}, -\vec{\nabla} \right)^2 + (E_n, \vec{P}_n)^2 + (E_n, -\vec{P}_n)^2 \right)$$

$$(E_n, \vec{P}_n) / 2 =$$

$$= \sum_{n=0}^{\infty} \left(-\vec{\nabla} \cdot (\vec{\nabla} X \vec{P}_n) - \vec{\nabla} \cdot \vec{\nabla} E_n - \right.$$

$$\vec{P}_n \cdot (\vec{P}_n X \vec{P}_n) - (\vec{P}_n \cdot \vec{P}_n) E_n + E_n^3 + \frac{\partial^2 E_n}{\partial t^2},$$

$$\vec{\nabla} X (\vec{\nabla} X \vec{P}_n) + \vec{\nabla} X (\vec{\nabla} E_n) + \vec{P}_n X (\vec{P}_n X \vec{P}_n) + (\vec{P}_n X \vec{P}_n) E_n - \vec{\nabla}$$

$$\left. (\vec{\nabla} \cdot \vec{P}_n) + \vec{P}_n E_n^2 - \vec{P}_n (\vec{P}_n \cdot \vec{P}_n) + \frac{\partial^2 \vec{P}_n}{\partial t^2} \right)$$

It takes some skilled staring to assure that this equation contains the Klein−Gordon equation along with vector identities.

■ Connection to the Maxwell Equations

If m=0, the quaternion operators of the Klein−Gordon equation simplifies to the operators used to generate the Maxwell equations in the Lorenz gauge. In the homogeneous case, the same operator acting on two different quaternions equals the same result. This implies that

$$(\varphi, \vec{A}) = \sum_{n=0}^{\infty} (E_n, \vec{P}_n)$$

Under this interpretation, a nonzero mass changes the wave equation into a simple harmonic oscillator. The simple relationship between the quaternion potential and the wave function may hold for the nonhomogeneous case as well.

■ Implications

The Klein−Gordon equation is customarily viewed as a scalar equation (due to the scalar D'Alembertian operator) and the Maxwell equations are a vector equation (due to the potential four vector). In this section, the quaternion operator that generated the Maxwell equations was used to generate the Klein−Gordon equation. This also created several vector identities which are usually not mentioned in this context. A quaternion differential equation is needed to perform the work of the Dirac equation, but since quaternion operators are a field, an operator that does the task must exist.

Time Reversal Transformations for Intervals

The following operator R for quaternions reverses time difference between two events:

$$(dt,\ d\vec{X}) \ \text{->}\ (-dt,\ d\vec{X}) = R\,(dt,\ d\vec{X})$$

The quaternion R exist because quaternions are a field.

The operator R will equal $(-dt, dX)(dt, dX)^{-1}$. The inverse of quaternion is the transpose over the square of the norm, which is the scalar term of the transpose of a quaternion times itself.

$$R = (-dt,\ d\vec{X})\,(dt,\ d\vec{X})^{-1} =$$
$$(-dt^2 + d\vec{X}.d\vec{X},\ 2\,dt\ d\vec{X})\ /\ (dt^2 + d\vec{X}.d\vec{X})$$

For any given event, the time−reversal operator R can be defined based on the above.

A criticism of this operator is that it is local, meaning it depends explicitly on spacetime. The two most important theories in physics, the standard model and general relativity, are also local, so this quality is consistent with those theories.

■ Classical Time Reversal

Examine the form of the quaternion which reverses time under two conditions. In the classical region, the change in time dt is much greater than space, dX. Calculate R in this limit to one order of magnitude in the ratio of dX/dt.

$$R = (-dt,\ d\vec{X})\,(dt,\ d\vec{X})^{-1} =$$
$$(-dt^2 + d\vec{X}.d\vec{X},\ 2\,dt\ d\vec{X})\ /\ (dt^2 + d\vec{X}.d\vec{X},\ 0)$$

$$\text{if}\ dt >> \left| dX \right|\ \text{then}\ R \approx \left(-1,\ 2\,\vec{\beta}\right)$$

The operator R is almost the negative identity, but the vector is non−zero, so it would not commute. In the classical limit, time reversal now depends on velocity, not the local position in spacetime, so classical time reversal is a global, not local operator.

■ Relativistic Time Reversal

For a relativistic interval involving one axis, the interval could be characterized by the following:

$$(T + \in,\ T,\ 0,\ 0)$$

Find out what quaternion is required to reverse time for this relativistic interval to first order in epsilon.

$$R = \left(\frac{T^2 - (T + \epsilon)^2}{T^2 + (T + \epsilon)^2}, \frac{2\,T\,(T + \epsilon)}{T^2 + (T + \epsilon)^2}, 0, 0 \right) =$$

$$\left(-\frac{\epsilon}{T} + O[\epsilon]^2, 1 + O[\epsilon]^2, 0, 0 \right)$$

This approaches q[−e/T, 1, 0, 0], almost a pure vector, a result distinct from the classical case. Again, in this limit, the transformation approaches a global transformation.

■ Implications

In special relativity, the interval between events is considered to be 4 vector are operated on by elements of the Lorentz group. The element of this group that reverses time has along its diagonal {−1, 1, 1, 1}, zeroes elsewhere. There is no dependence on relative velocity or any local information. Therefore special relativity predicts the operation of time reversal should be indistinguishable for classical and relativistic intervals. Yet classically, the boundary condition of time reversal appears to involve entropy. For relativistic interactions, time reversal involves antiparticles.

In this section, a time reversal quaternion operator has been derived and shown to work. Time reversal for classical and relativistic intervals have distinct limits, but these transformations have not yet been tied explicitly to laws of physics.

Gravity

This is a current area of active research for me. I have two distinct approaches. The first uses scalars, vectors, tensors, and the Christoffel symbol to characterize a dynamic metric. The work attempts to unify gravity and EM in a rank 1 field theory using an asymmetric field strength tensor in the Lagrange density. It has three testable predictions. Light should be bent more at second order parameterized post−Newtonian accuracy. In a gravity wave, the mode of emission will not be transverse (that mode of emission is done by light). There is also an new constant velocity solution for gravity that may eliminate the need for dark matter for the velocity profile of thin disk galaxies.

The approach to gravity that uses quaternions makes the same three experimental predictions. The difference is that the Christoffel symbol is not used. It will require the efforts of someone with more technical math skills than I possess to determine if this approach is valid. There are aspects of the quaternion approach I prefer, but at this time I am promoting the form that uses the Christoffel since the odds of it being accepted are higher.

Gravity by Analogy to EM

■ Abstract

A Lagrange density for gravity is proposed based on a strict analogy to the classical Lagrangian for electromagnetism. The field equations are a four−dimensional wave equation. The classic field equations contain both the Maxwell equations and Newton's field equations under certain conditions. The four−dimensional wave equation has been quantized before. The scalar and longitudinal modes of emission are interpreted as spin 2 gravitons, so they can do the work of gravity. If gravitational waves are detected, this proposal predicts scalar or longitudinal polarization. How the proposal integrates with the standard model Lagrangian is worked out.

A force equation is written based on the same strict analogy to the relativistic Lorentz force of electromagnetism. For geodesic motion, the cause of the curvature is due entirely to the gravitational and electric potentials. This is a new type of statement about curvature. A specific, normalized, weak−field potential is investigated. Analysis of small perturbations yields changes in the potential that depend on an inverse distance squared. By breaking spacetime symmetry, Newton's law of gravity results. By using the chain rule, a stable, constant−velocity solution is apparent, which may yield insight to the rotation profile of galaxies and early big bang cosmology, since both require stable, constant−velocity solutions. If spacetime symmetry is preserved, the second−order differential equations can be solved exactly. Eliminating the constants and rearrange terms generates an equation that has the form of a metric equation. The Taylor series expansion of the metric equation is identical to the Schwarzschild metric to parameterized−post−Newtonian accuracy. The Taylor series for the two metrics differ for higher order terms and may be tested experimentally.

■ Introduction

The goal of this paper is to create one mathematical structure for gravity and electromagnetism that can be quantized. The difference between gravity and electromagnetism is the oldest core problem facing physics, going back to the first studies of electromagnetism in the seventeenth century. Gravity was the first inverse square law, discovered by Isaac Newton. After twenty years of effort, he was able to show that inside a hollow massive shell, the gravitational field would be zero. Ben Franklin, in his studies of electricity, demonstrated a similar property using a conducting cup. Joseph Priestly realized this meant that the electrostatic force was governed by an inverse square law just like gravity. Coulomb got the credit for the electrostatic force law modeled on Newton's law of gravity.

Over a hundred years later, Einstein started from the tensor formalism of electromagnetism on the road to general relativity. Instead of an antisymmetric field strength tensor, Einstein used a symmetric tensor because the metric tensor is symmetric. There is a precedence for transforming mathematical structures between gravity and electromagnetism.

The process of transforming mathematical structures from electromagnetism to gravity will be continued. Specifically, the gravitational analog to the classic electromagnetic Lagrange density will be written. There are several consequence of this simple procedure. The Lagrangian contains both terms with a connection and the Fermi Lagrangian of electromagnetism. This makes it reasonable to suppose the Lagrangian can describe both a dynamic geometry required for gravity and the

Maxwell equations for electrodynamics. The gravitational field equations are analogues to Gauss' and Ampere's laws, and contain the Newton's gravitational field equation. These field equations are not second rank like those used in general relativity. It must be stressed that the field strength tensor is a second order symmetric tensor, so this does not conflict with proofs that at least a symmetric second rank tensor is required to completely describe spacetime curvature. The Maxwell equations result if the mass current and gravitational field are zero. The field equations have been quantized before, but new interpretations will flow from the unification effort. A link to the Lagrangian of the standard model will be detailed.

A weak static gravitational field in a vacuum will be studied using standard modern methods: normalizing the potential and looking at perturbations. The potential will be plugged into a gravitational force equation analogous to the Lorentz force equation of electromagnetism. The force equation leads to a geodesic equation where the potential causes the curvature, something which is missing from general relativity. Newton's law of gravity is apparent if spacetime symmetry is broken. A new class of solutions emerges for the gravitational source where velocity is constant, but the distribution of mass varies with distance. This may provide new ways to look at problems with the rotation profiles of disk galaxies and big bang cosmology. If spacetime symmetry is preserved, solving the force equation and eliminating the constants creates a metric equation similar to the Schwarzschild metric. The metrics are equivalent to first−order parameterized post−Newtonian accuracy. Therefore the metric will past all weak field tests. The coefficients are different to second−order, so the proposal can be verify or rejected experimentally.

■ Lagrangians

The classic electromagnetic Lagrangian density has three terms: one for kinetic energy, one for a moving change, and a third for the antisymmetric second rank field strength tensor:

$$\mathcal{L}_{\mathrm{EM}} = -\frac{\rho}{\gamma} - \frac{1}{c} J^{\mu} A_{\mu} - \frac{1}{4 c^2} (\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu}) (\partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu})$$

An analogous Lagrangian for gravity would also contain these three components, but three changes are required. First, gravity depends on mass, not electric charge, so where there is an electrical charge −q, an inertial mass +m will be substituted. The change in sign is required so that like charges attract for gravity. Mass does not have the same units as electric charge, so mass will have to be multiplied by the square root of Newton's gravitational constant G to keep the units identical. Second, because gravity effects metrics which are symmetric, the source of gravity must also be symmetric. The minus sign that makes the electromagnetic field strength tensor antisymmetric will be made positive. Third, in order that symmetric object transforms like a tensor requires a replacement of the exterior derivative with a covariant derivative:

$$\mathcal{L}_{\mathrm{G}} = -\frac{\rho}{\gamma} + \frac{1}{c} J_m{}^{\mu} A_{\mu} - \frac{1}{4 c^2} (\nabla_{\mu} A^{\nu} + \nabla_{\nu} A^{\mu}) (\nabla^{\mu} A_{\nu} + \nabla^{\nu} A_{\mu})$$

The total Lagrangian will be a merger of these two which only apply if the other force is not in effect. The kinetic energy term is the same as either Lagrangian separately. The moving charge term is a sum. Without loss of generality, the regular derivatives in the electromagnetic Lagrangian (Eq. L_EM) can be written as covariant derivatives. This leads to the unified Lagrangian for gravity and electromagnetism:

$$\mathcal{L}_{\mathrm{GEM}} = -\frac{\rho}{\gamma} - (J_q{}^{\mu} - J_m{}^{\mu}) A_{\mu} - \frac{1}{2 c^2} \nabla_{\mu} A^{\nu} \nabla^{\mu} A_{\nu}$$

$$= -\frac{\rho}{\gamma} - \frac{q - \sqrt{G}\, m}{c^2\, V}\, (\phi - \vec{A} \cdot \vec{v}) -$$

$$\frac{1}{2\, c^2}\, \partial_\mu A^\nu\, \partial^\mu A_\nu - \frac{1}{2\, c^2}\, \Gamma^{\omega\,\mu}{}_\nu\, (\Gamma_{\rho\mu}{}^\nu\, A^\rho\, A_\omega + 2\, \partial_\mu A^\nu\, A_\omega)$$

The kinetic energy term is for one particle experiencing both gravity and electromagnetism. The Fermi Lagrangian of electromagnetism is a subset. This establishes a link to electromagnetism. The Christoffel symbols (or connection coefficients) represent derivatives of metrics. Because a dynamic metric is part of the Lagrangian, this Lagrangian could describe the dynamics of the metric, which is a central accomplishment of general relativity. The potential to do both gravity and electromagnetism is here.

In local covariant coordinates, the connect is zero, which leads to a simpler expression of the Lagrangian:

$$\mathcal{L}_{GEM} = -\frac{m}{V}\sqrt{1 - \left(\frac{\partial \vec{R}}{\partial t}\right)^2} - \frac{q - \sqrt{G}\, m}{c^2\, V}\, (\phi - \vec{A} \cdot \vec{v}) -$$

$$\frac{1}{2\, c^2}\left(\left(\frac{\partial \phi}{\partial t}\right)^2 - (\vec{\nabla}\phi)^2 - \left(\frac{\partial \vec{A}}{\partial t}\right)^2 + (\nabla \vec{A})^2\right)$$

This is almost identical to working with the classical electromagnetic field equation by choosing the Lorenz gauge, the difference being the inclusion of a mass term. Because the gauge was not fixed, there is more freedom for this Lagrangian, which is required if this field equation does more than just electromagnetism.

■ Classical Field Equations

The field equations can be found by applying the Euler−Lagrange equations to the Lagrange density (assuming the connection is zero for simplicity):

$$\Box^2 A^\mu = J_q{}^\mu - J_m{}^\mu$$

The fields are expressed in terms of the potential. The symmetric and antisymmetric field strength tensors are very similar, differing only in the sign of A^v;u. The classical fields required to represent the field strength tensors should also be similar. There is a symmetric analog to the electric E and B fields: To make a connection to the classical fields of gravity and electromagnetism, use the following substitutions:

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} + c\, \vec{\nabla}\phi$$

$$\vec{e} = \frac{\partial \vec{A}}{\partial t} + c\, \vec{\nabla}\phi$$

$$\vec{B} = -c\left(\frac{\partial}{\partial y} - \frac{\partial}{\partial z},\ \frac{\partial}{\partial x} - \frac{\partial}{\partial z},\ \frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\vec{A} = -c\, \vec{\nabla} \times \vec{A}$$

$$\vec{b} = c\left(-\frac{\partial}{\partial y} - \frac{\partial}{\partial z},\ -\frac{\partial}{\partial x} - \frac{\partial}{\partial z},\ -\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\vec{A} \equiv \vec{\nabla} \underline{\times} \vec{A}$$

The symmetric curl as defined above has all the same differential operators, but all the signs are negative, so it is easier to remember. The symmetric field strength tensor has four more components that lie along the diagonal. Define a field g to represent the diagonal elements:

$$g = \left(\frac{\partial \phi}{\partial t}, \ c\,\frac{\partial A_x}{\partial x}, \ c\,\frac{\partial A_y}{\partial y}, \ c\,\frac{\partial A_z}{\partial z} \right) = \nabla_\mu A^\mu$$

The diagonal of the field strength tensor $\nabla_\mu A^n$ is g. The first row and column of the asymmetric field strength tensor is the sum of the electric field E and its symmetric analog e. The rest of the off−diagonal terms are the sum of the magnetic field B and its symmetric analog b. If the trace of field strength tensor is zero, then the equations are in the Lorentz gauge.

Substitute the classical fields into the field equations, starting with the scalar field equation. Important technical note: all the ∇ operators used for the divergences and curls are contravariant, that means they bring in an extra minus sign to these expressions for the field equations of the form $\nabla^\mu (\nabla_\mu A^\nu) = J^\nu$:

$$\rho_q - \rho_m = \frac{1}{c}\,\frac{\partial^2 \phi}{\partial t^2} - c\,\vec{\nabla}^2 \phi$$

$$= \frac{c}{2}\,(\vec{\nabla} \cdot \vec{E} + \vec{\nabla} \cdot \vec{e}) + \frac{\partial g^0}{\partial t}$$

This equation combines Gauss' law and analogous equation for gravity. The two equations are unified, but under certain physical conditions, can be isolated. A relativistic form of the Newtonian gravitational field equation can be seen with the following constraints:

$$-\rho_m = -c\,\vec{\square}^2 \phi$$

$$\text{iff } \rho_q = 0, \quad \text{and} \quad \frac{\partial \vec{A}}{\partial t} = -c\,\vec{\nabla}\phi$$

This equation should be consistent with special relativity without modification. The classical Newtonian field equation arises from these physical constraints:

$$\rho_m = c^2\,\vec{\nabla}^2 \phi$$

$$\text{iff } \rho_q = 0, \ \frac{\partial \vec{A}}{\partial t} = -c\,\vec{\nabla}\phi \ \text{ and } \ \frac{\partial g^0}{\partial t} = 0$$

Every aspect of classical Newtonian gravity can be represented by this proposal under these constraints.

Gauss' law appears under the following conditions:

$$\rho_q = \frac{1}{c}\,\frac{\partial^2 \phi}{\partial t^2} - c\,\vec{\nabla}^2 \phi$$

$$\text{iff } \rho_m = 0, \quad \text{and} \quad \frac{\partial \vec{A}}{\partial t} = c\,\vec{\nabla}\phi$$

Repeat the exercise for the vector equation.

$$\vec{J}_q - \vec{J}_m = \frac{1}{c}\frac{\partial^2 A}{\partial t^2} - c\,\nabla^2 \vec{A}$$

$$= \frac{1}{2}\left(-\frac{\partial \vec{E}}{\partial t} + c\,\vec{\nabla} \times \vec{B} + \frac{\partial \vec{e}}{\partial t} - c\,\vec{\nabla} \underline{\times}\, \vec{b}\right) + 2\,c\,\vec{\nabla} g^u$$

This has Ampere’s law and a symmetric analog for Ampere’s law for gravity.

This proposal for classical gravitational and electromagnetic field equations is expressed with tensors of rank one (vectors). Einstein’s field equations are second rank. Therefore the two approaches are fundamentally different. One must remember that although the field equations are rank one, the field strength tensor is second rank.

With no gravitational field, the Maxwell source equations result. The homogeneous Maxwell equations are vector identities with these choices of maps to the potentials, and are unaffected by the proposal.

■ Canonical Quantization

The classical electromagnetic Lagrangian cannot be quantized. One way to realize this is to consider the generalized 4−momentum:

$$\pi^\mu = h\sqrt{G}\,\frac{\partial \mathcal{L}_{EM}}{\partial\left(\frac{\partial A^\mu}{c\,\partial t}\right)} = -F^{\mu 0}$$

Unfortunately, the energy component of the moment operator is zero. The commutator $[A^0, \pi^0]$ will equal zero, and cannot be quantized. The momentum for the unified Lagrangian of gravity and electromagnetism does not suffer from this problem:

$$\pi^\mu = h\sqrt{G}\,\frac{\partial A^\mu}{c\,\partial t}$$

When expressed with operators, the commutator $[A^0, \pi^0]$ will not be zero, so the field can be quantized. If the connection is zero, L_GEM generates the same field equations as the classical electromagnetic Lagrangian with the choice of the Lorenz gauge. That field has been quantized before, first by Gupta and Bleuler (S. N. Gupta, Proc. Phys. Soc. London, 63:681−691, 1950). They determined that there were four modes of transmission: two transverse, one scalar, and one transverse mode. The interpretation of these modes appears internally inconsistent to this author. They discuss "scalar photons", but photons as the quanta of electric and magnetic fields must transform as a vector, not a scalar. They introduce a supplemental condition solely to make the scalar and longitudinal modes virtual. Yet there is no need to make a nonsense particle virtual.

The field in this proposal must represent both gravity and electromagnetism. The two transverse modes are photons that do all the work of electromagnetism. The symmetric second−rank field strength tensor cannot be represented by a photon because photons transform differently than a symmetric tensor. Whatever particle does the work must travel at the speed of light like the transverse modes of transmissions of the field. These constraints dictate that the scalar and transverse modes of transmission for this proposal are gravitons.

There are efforts underway to detect the transverse gravitational waves predicted by general relativity. This proposal predicts the polarity of a gravitational wave will be either scalar or longitudinal, not transverse, because those are the modes of transmission. The detection of the first gravitational wave polarization will mark either success or failure of this unified field theory.

■ Integration with the Standard Model

The standard model does not in an obvious way deal with curved spacetime. A more explicit connection will be attempted by condensing the unitary aspects of the symmetries U(1), SU(2), and SU(3) with the 4−vectors and a curved metric. Start with the standard model Lagrangian:

$$\mathcal{L}_{SM} = \bar{\Psi}\,\gamma^{\mu}\,D_{\mu}\,\Psi$$

where

$$D_{\mu} = \partial_{\mu} - i g_{EM}\, Y A_{\mu} - i g_{weak}\,\frac{\tau^{a}}{2}\,W^{a}_{\mu} - i g_{strong}\,\frac{\lambda^{b}}{2}\,G^{b}_{\mu}$$

The electromagnetic potential A_mu is a complex−valued 4−vector. The only way to form a scalar with a 4−vector is to use a metric. Since it is complex−valued, use the conjugate like so:

$$A^{\mu}\,A^{\nu *}\,g_{\mu\nu} = |A_0|^2 - |A_1|^2 - |A_2|^2 - |A_3|^2$$

Use the parity operator to flip the sign of the spatial part of a 4−vector:

$$A^{\mu}\,A^{\nu * p}\,g_{\mu\nu} = |A_0|^2 + |A_1|^2 + |A_2|^2 + |A_3|^2$$

Normalize the potential:

$$\frac{A^{\mu}}{|A|}\,\frac{A^{\nu * p}}{|A|}\,g_{\mu\nu} = 1$$

From this, it can be concluded that the normalized 4−vector is an element of the symmetry group U(1) if the multiplication operator is the metric combined with the parity and conjugate operators. One does not need the Y in standard model Lagrangian, so this simplifies things. The same logic applies to the 4−vector potentials for the weak and the strong forces which happen to have internal symmetries.

In curved spacetime, the previous equation will not equal one. Mass breaks U(1), SU(2), and SU(3) symmetry, but does so in a precise way (meaning one can calculate what the previous equation should equal). There is no need for the Higgs mechanism to give particles mass while preserving U(1)xSU(2)xSU(3) symmetry, so this proposal predicts no Higgs particle will be found.

■ Forces

The Lorentz Force of electromagnetism involves charge, velocity and the anti−symmetric field strength tensor:

$$F^{\mu}_{EM} = q\,\frac{U_{\nu}}{c}\,(\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu})$$

Form an analogous force for gravity using the same substitutions as before:

$$F_{G\,\mu} = -\sqrt{G}\,m\,\frac{U_{\nu}}{c}\,(\nabla_{\mu} A^{\nu} + \nabla_{\nu} A^{\mu})$$

The gravitational force and the electromagnetic force behave differently under charge inversion. If the mass changes signs, then both side flip signs, so nothing has really changed. If electric change changes signs, the change in momentum will not change signs. The different behavior under charge inversion may explain why gravitational force is unidirectional, but electrical forces can attract or repulse.

The total force is a combination of the two:

$$F_{GEM\,\mu} = (J_{q\,\nu} - J_{m\,\nu})\, \nabla_\mu A^\nu - (J_{q\,\nu} + J_{m\,\nu})\, \nabla_\nu A^\mu$$

If q >> $\sqrt{G}\, m$, the equation approaches the form of the Lorentz force law of electromagnetism. If the force is zero, the equation has the form of a Killing's equation, which is used to determine the isometries of a metric. Geodesics are defined by examining the left−hand side of F_GEM:

$$\frac{\partial m\, U^\mu}{\partial \tau} = m \frac{\partial U^\mu}{\partial \tau} + U^\mu \frac{\partial m}{\partial \tau} = 0$$

Assume dm/dtau = 0. Apply the chain rule, and then the definition of a covariant derivative to form a geodesic equation:

$$0 = m \frac{\partial^2 x^\mu}{\partial \tau^2} + \frac{m}{c} \Gamma^\mu{}_{\nu\omega} U^\nu U^\omega$$

This equation says that if there is no force, all the acceleration seen in spacetime is due to spacetime curvature, the Christoffel symbol. The covariant derivatives on the right side of F_gEM can also be expanded:

$$0 = \left(q - \sqrt{G}\, m\right) \frac{\partial x_\nu}{c\, \partial \tau} \partial_\mu A^\nu - \left(q + \sqrt{G}\, m\right) \frac{\partial x_\nu}{c\, \partial \tau} \partial_\nu A^\mu - \frac{2}{c} m \Gamma_\omega{}^{\mu\nu} U_\nu U^\omega$$

This equation says that spacetime curvature is caused by the change in the potential if there is no external force. This is a novel statement. In general relativity, one compares two geodesics, and based on an analysis of the tidal forces between the geodesics, determines the curvature. The unified geodesic equation asserts that the curvature can be calculated directly from the potential. Notice that this equation contains terms linked to a mass m and a charge q, so the geodesic equation applies to electromagnetism as well as gravity.

■ Gravitational Force for a Weak Field

The total unified force law is relevant to physics because it contains the Lorentz force law of electromagnetism. It must be established that the terms coupled to the mass m are connected to what is known about gravity.

The next task is to find a solution to the unified field equations, and then put the solutions into the force equation. The Poisson field equation of classical Newtonian gravity can be solved by a 1/R potential. The potential has a point singularity where R = 0. The unified field equations are relativistic, so time must also be incorporated. A 1/distance potential does not solve the field equations in four dimensions. In local covariant coordinates where the connection is zero, the potential A_mu = (1/sigma^2, 0) solves the field equations, where sigma squared is the Lorentz invariant distance, or the negative of the square of the Lorentz invariant interval tau. Distance is used instead of the interval because classical gravity depends on distance, not time. The idea is to consider the time contribution to be very small relative to the distance. Such a potential has as a singularity that is the entire lightcone, where sigma ^2 = 0. This singularity may not be problematic because massless particles are described by the Maxwell equations, but that hope will required a detailed study.

Gravity is a weak effect. It is common in quantum mechanics to normalize to one and study perturbations of weak fields, an approach that will be followed here. Normalizing means there are small steps will be away from one. Only first order terms will be kept. Here is the normalized potential with a linear perturbation:

$$A^{\mu} = \left(\frac{\sqrt{G}\,h}{c^2\,\sigma^2}, \vec{0}\right) \longrightarrow$$

$$\left(c \Big/ \left(\sqrt{G}\left(\left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,x\right)^2 + \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,y\right)^2 + \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,z\right)^2 - \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,t\right)^2\right)\right), \vec{0}\right)$$

This potential solves the 4D wave equation because the shift by the one over root two factor and the rescaling by the spring constant k over sigma square do not effect the differential equation. One interesting aspect is the shift of units from one that depends on h − suggesting quantum mechanics − to the normalized perturbation which appears to be classical because there is no h.

Take the derivative with respect to t, x, y, and z:

$$\frac{\partial\phi}{\partial t} = \frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\left(k^2\right)$$

$$c\,\frac{\partial\phi}{\partial x} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\left(k^2\right)$$

$$c\,\frac{\partial\phi}{\partial y} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\left(k^2\right)$$

$$c\,\frac{\partial\phi}{\partial z} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\left(k^2\right)$$

The change in the potential is a function of a spring constant k over sigma squared. The classical Newtonian dependence on distance is an inverse square, so this is promising. One problem is that a potential that applies exclusively to gravity is sought, yet the non−zero gradient of ϕ indicates an electric field. The sign of the spring constant k does not effect the solution to the 4D wave field equations but does change the derivative of the potential. A potential that only has derivatives along the diagonal of the field strength tensor can be constructed from two potentials that differ by spring constants that either constructively interfere to create non−zero derivatives, or destructively interfere to eliminate derivatives.

$$\text{diagonal SHO } A^{\mu} =$$

$$\frac{c}{\sqrt{G}}\left(1 \Big/ \left(\left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,x\right)^2 + \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,y\right)^2 + \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,z\right)^2 - \left(\frac{1}{\sqrt{2}} + \frac{k}{\sigma^2}\,t\right)^2\right) + \right.$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,t\right)^2\right)\quad,$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,t\right)^2\right)+$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,t\right)^2\right)\quad,$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,t\right)^2\right)+$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,t\right)^2\right)\quad,$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,t\right)^2\right)+$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}\,z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}\,t\right)^2\right)\Big)$$

Take the contravariant derivative of this potential, keeping only the terms to first order in the spring constant k.

$$\nabla_\mu A^\nu = \frac{c^2}{\sqrt{G}} \frac{k}{\sigma^2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

All this work to get a multiple of the identity matrix! Plug this into the gravitational force equation:

$$F_g^\mu = m\,c \left(-\frac{k}{\sigma^2} \frac{\partial t}{\partial \tau},\ \frac{k}{\sigma^2} \frac{\partial \vec{R}}{\partial \tau} \right)$$

This is a relativistic force law for a weak gravitational field for the inverse interval squared diagonal potential. When spacetime symmetry is broken, this equation will lead to Newton's law of gravity in the next section. If spacetime symmetry is maintained, then solving the force equation and eliminating the constants yields a metric equation for gravity.

■ Newton's Law of Gravity and More

Several assumptions need to be made to apply the weak gravitational force equation to a classical gravitational system. First, assume that the spring constant is due to the source mass, k = GM. Second, assume that the field is static, so that $\sigma^2 = R^2 - c^2\, t^2 \simeq R^2$. In the approximation, it does not depend on time.

Newtonian spacetime is different from Minkowski spacetime because the speed of light is infinite. Spacetime symmetry must be broken. A question arises about how to do this in a formal mathematical sense. The Minkowski interval tau is a consequence of the relationship between time t and space R. The functional relationship between time and space must be severed. By the static field approximation, there is a distance R which is the same magnitude as the interval tau. If the interval tau is replaced by the scalar distance R, then that will sever the functional relationship between time and space:

$$\left(\frac{\partial t}{\partial \tau},\ \frac{\partial \vec{R}}{c\,\partial \tau} \right) \longrightarrow \left(\frac{c\,\partial t}{\partial\,|\,R\,|},\ \frac{\partial \vec{R}}{\partial\,|\,R\,|} \right) = (0,\ \hat{R})$$

Plug these three assumptions into force equation:

$$F_g^\mu = \left(0,\ -\frac{G\,M\,m}{R^2}\ \hat{R} \right)$$

This is not quite Newton's gravitational force law. The reason is that one must consider the left−hand side of the force equation carefully. According to the chain rule:

$$\frac{\partial m\,U^\mu}{\partial \tau} = m\,\frac{\partial U^\mu}{\partial \tau} + U^\mu\,\frac{\partial m}{\partial \tau}$$

An open question is how should spacetime symmetry be broken for the derivatives with respect to the interval tau? An interval is composed of both changes in time and space. For the acceleration term, if the interval is only about time, then one gets back Newtonian acceleration. For logical consistency, one might be tempted to also substitute time in the dm/dtau term. However, the system is presumed to be static, so this would necessarily be zero. If this derivative is to have any chance at being non−zero, it would have to be with respect to the absolute value of R as has been done earlier in the derivation. So the classical force law should look like so:

$$m \frac{\partial^2 \vec{R}}{\partial t^2} + c \frac{\partial \vec{R}}{\partial t} \frac{\partial m}{\partial |R|} = -\frac{G M m}{R^2} (\hat{R} + \hat{V})$$

This is Newton's law of gravity working along a new direction, that for the velocity vector. This new effect will necessarily be small due to the constant c in the term.

For a point source, the dm/d|R| term will not make a contribution, and one gets Newton's law of gravity. It is only if the inertial mass is distributed over space like for the big bang or galaxies will the term come into play. If the velocity is constant, then the acceleration is zero. The equation describes the distribution of the inertial mass m that makes up the total gravitational source mass M. The solution to the force equation when there is no acceleration is a stable exponential. Big bang cosmology has two problems: all matter is traveling at exactly the same speed even though it is not possible for them to communicate (the horizon problem), and the model require high levels of precision on initial conditions to avoid collapse (the flatness problem). [A. H. Guth, Phys. Rev. D., 23:347−356, 1981] The force equation has a stable, constant velocity solution which may resolve both problems of the big bang without the inflation hypothesis. Their is also a problem with the rotation profile of thin disk galaxies.[S. M. Kent, Astron. J., 91:1301−1327, 1986; S. M. Kent, Astron. J., 93:816−832, 1987] Once the maximum velocity is reached, the velocity stays constant. It has been shown that galaxies should not be stable at all.[A. Toomre, Astrophys. J., 139:1217, 1964] Both problems may again be resolved with stable constant velocity solutions. Numerical approaches on the above equation should be conducted.

■ A Metric Equation

The weak gravitational force equation is two second−order differential equations. The equation can be simplified to a set of first−order differential equations by substituting (U^0, U) = (c dt/dtau, dR/dtau)

$$\frac{\partial U^0}{\partial \tau} - \frac{k}{\tau^2} U^0 = 0$$

$$\frac{\partial \vec{U}}{\partial \tau} + \frac{k}{\tau^2} \vec{U} = 0$$

The solution involves exponentials:

$$U^\mu = \left(v e^{-\frac{k}{\tau}}, \vec{V} e^{\frac{k}{\tau}} \right)$$

For flat spacetime, U^mu = (v, V). The constraint on relativistic velocities in flat spacetime is:

$$U^\mu U_\mu = \frac{c^2 dt^2 - dR^2}{d\tau^2} = c^2 = v^2 - \vec{V} \cdot \vec{V}$$

Solve for the constants, and plug back into the constraint, multiplying through by dtau ^2.

$$d\tau^2 = e^{-2 \frac{k}{\tau}} dt^2 - e^{2 \frac{k}{\tau}} \frac{dR^2}{c^2}$$

Make the same two assumptions as before: the spring constant is due to the gravitational source, k = GM/c^2, and the field is static, so $\tau^2 = R^2 / c^2 - t^2 \simeq R^2 / c^2$. There is one more degree of freedom, because the radius R could either be positive or negative. To make the metric consistent with experiment, choose the negative root:

$$d\tau^2 = e^{-2\frac{GM}{c^2 R}}\, dt^2 - e^{2\frac{GM}{c^2 R}}\, dR^2$$

This equation has the form of a metric equation. Perform a Taylor series expansion to second order in GM/c^2R:

$$d\tau^2 = \left(1 - 2\frac{GM}{c^2 R} + 2\left(\frac{GM}{c^2 R}\right)^2\right) dt^2 - \left(1 + 2\frac{GM}{c^2 R} + 2\left(\frac{GM}{c^2 R}\right)^2\right) dR^2$$

If one compares this metric to the Schwarzschild metric in isotropic coordinates to parameterized post−Newtonian accuracy, the coefficients are identical. For that reason, this metric is consistent with all experimental tests weak field tests of general relativity. [C. M. Will, "Theory and experiment in gravitational physics: Revised edition", Cambridge University Press, 1993.]

For higher order terms of the Taylor series expansion, the two metric will predict different coefficients. The validity of this proposal can thus be tested experimentally. It will require a great deal of effort and skill to conduct such experiments, since many physical phenomena will have to be accounted for (an example: the quadrupole moment of the Sun for solar tests). According to personal communication with Prof. Clifford Will − a leading authority in experimental tests of gravity theories − no experiments are being planned to monitor second order PPN coefficients at this time.

■ Conclusion

Using a nineteenth century approach, an effort to unify physics from the twentieth century has been attempted. The description of geodesics by general relativity is not complete because it does not explicitly show how the potential source causes curvature. A dynamic metric equation is found but it uses a simpler set of field equations (a rank one tensor instead of two). In the standard model as elsewhere, combining two 4−vectors requires a metric. By normalizing the 4−vectors, the unitary aspect of the standard model can be self−evident.

This theory makes three testable predictions, two subtle, one not. First, the polarity of gravitational waves will be scalar or longitudinal, not transverse as predicted by general relativity. Second, if gravitation effects are measured to secondary parameterized post−Newtonian accuracy, the coefficients for the metric derived here are different from the Schwarzschild metric in isotropic coordinates. Such an experiment will be quite difficult to do. The third test is to see if the complete relativistic force equation matches all the data for a thin spiral galaxy. It is this test which should be investigated first.

Einstein's Vision I: Classical Unified Field Equations

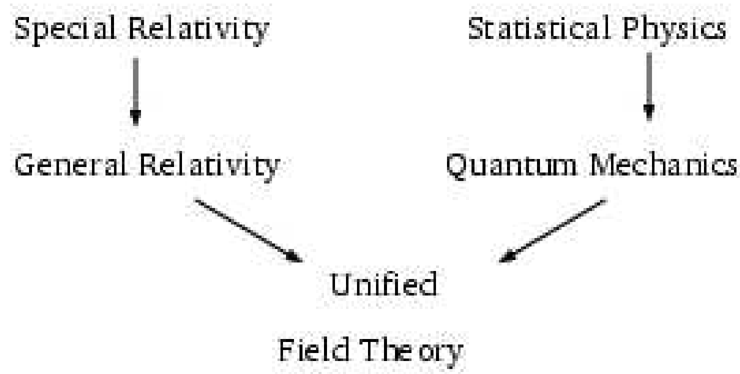
■ Abstract

The equations governing gravity and electromagnetism show both profound similarities and unambiguous differences. Albert Einstein worked to unify gravity and electromagnetism, mainly by trying to generalize Riemannian geometry. Hamilton's quaternions are a 4−dimensional topological algebraic field related to the real and complex numbers equipped with a static Euclidean 4−basis. Riemannian quaternions as defined herein explicitly allow for dynamic changes in the basis vectors. The equivalence principle of general relativity which applies only to mass is generalized because for any Riemannian quaternion differential equation, the chain rule means a change could be caused by the potential and/or the basis vectors. The Maxwell equations are generated using quaternion potentials and operators. Unfortunately, the algebra is complicated. The unified force field proposed is modeled on a simplification of the electromagnetic field strength tensor, being formed by a quaternion differential operator acting on a potential, Box* A*. This generates an even, antisymmetric−matrix field strength quaternion for electricity and an odd, antisymmetric−matrix field strength quaternion for magnetism, where the even field conserves its sign if the order of the differential and the potential are reversed unlike the odd field. Gauge symmetry is broken for massive particles by the even, symmetric−matrix term, which is interpreted as being due to gravity. In tensor analysis, a differential operator acting on the field strength tensor creates the Maxwell equations. The unified field equations for an isolated source are generated by acting on the unified force field with an additional differential operator, Box* Box* A* = 4 pi J*. This contains a quaternion representation of the Maxwell equations, a classical link to the quantum Aharonov−Bohm effect, and dynamic field equations for gravity. Vacuum and zero net current solutions to the unified field equations are discussed. The field equations conserve both electric charge density and mass density. Under a Lorentz transformation, the gravitational and electromagnetic fields are Lorentz invariant and Lorentz covariant respectively, but there are residual terms whose meaning is not clear presently. An additional constraint is required for gauge transformations of a massive field. (PACS:12.10.−g)

■ Einstein's Vision Using Quaternions

Three of the four known forces in physics have been unified via the standard model: the electromagnetic, the weak, and the strong forces. The holdout remains gravity, the first force characterized mathematically by Isaac Newton. The parallels between gravity and electromagnetism are evident. Newton's law of gravity and Coulomb's law are inverse square laws. Both forces can be attractive, but Coulomb's law can also be a repulsive force. A long−standing goal of modern physics is to explain the similarities and differences between gravity and electromagnetism.

Albert Einstein had a specific idea for how to formulate an acceptable unified field theory (see Fig. 1, taken from A. Pais, "Subtle is the Lord..." the science and life of Albert Einstein", Claredon Pres, 1982).



One unusual aspect of Einstein's view was that he believed the unified field would lead to a new foundation for quantum mechanics, an idea which is not shared by most of today's thinkers (S. Weinberg, "Dreams of a final theory," Pantheon Books, New York, 1992). Most of Einstein's efforts over 40 years were directed in a search to generalize Riemannian differential geometry in four dimensions.

To a degree which has pleasantly surprised the author, Einstein's vision to unify gravity and electromagnetism has been followed. The construction of a new 4−dimensional geometry is dictated by insights garnered from physics. Events in spacetime are composed of a scalar for time and a 3−vector for space. The four−dimensional topological algebraic field of quaternions has the same structure, so quaternions will be the starting point of this effort.

Laws of physics are expressed in a coordinate−independent way. The sum or difference of two quaternions can only be defined if the two quaternions in question share the same 4−basis. Riemannian quaternions make coordinate−independence explicit. In special relativity, regions in spacetime are delimited by the light cone, where the net change in 3−space is equal to the net change in time. The parity between changes in 3−space and time is constructed into the definition of a Riemannian quaternion. In general relativity, the field equations make the metric a dynamic variable. The basis vectors of Riemannian quaternions can be dynamic, so the metric can be dynamic. The dynamic nature of the basis vectors leads to the general equivalence principle, whereby any law, even those in electromagnetism, can be the result of a change in reference frame.

Physical laws are the result of simple Riemannian quaternion differential equations. First−order Riemannian quaternion differential equations create force fields for gravity, electricity, and magnetism. Second−order differential equations create dynamic field equations for gravity, the Maxwell equations for electromagnetism, and a classical counterpart to the Aharanov−Bohm effect of quantum mechanics. Third−order differential equations create conservation laws. Homogeneous solutions to the second order differential equations are related to gauge symmetry.

The second paper in this series of three investigates a unified force law, with a focus on a particular solution which may eliminate the need for dark matter to explain the mass distribution and velocity profile for spiral galaxies. The third paper develops a new approach to quaternion analysis. The equations of the first two papers are recast with the new definition of a quaternion derivative, resulting in a quantum unified field and force theory.

■ Events in Spacetime and Quaternions

An event in spacetime is considered by the author as the fundamental form of information in physics. Events have structure. There are four degrees of freedom divided into two dissimilar parts: time is a scalar, and space is a 3−vector. This structure should be reflected in all the mathematics used to describe patterns of events. For this reason, this paper focuses exclusively on quaternions, the 4−dimensional number where the terms scalar and vector where first used.

Hamilton's quaternions, along with the far better know real and complex numbers, can be added, subtracted, multiplied, and divided. Technically, these three numbers are the only finite−dimensional, associative, topological, algebraic fields, up to an isomorphism (L. S. Pontryagin, "Topological groups", translated from the Russian by Emma Lehmer, Princeton University Press, 1939). Properties of these numbers are summarized in the table below by dimension, if totally ordered, and if multiplication commutes:

Number	Dimensions	Totally Ordered	Commutative
Real	1	Yes	Yes
Complex	2	No	Yes
Quaternions	4	No	No

Hamilton's quaternions have a Euclidean 4−basis composed of 1, i, j, and k. The rules of multiplication were inspired by those for complex numbers: 1^2=1, i^2=j^2=k^2=ijk=−1. Quaternions also have a real 4x4 matrix representation:

$$q(t, x, y, z) = \begin{pmatrix} t & -x & -y & -z \\ x & t & -z & y \\ y & z & t & -x \\ z & -y & x & t \end{pmatrix}$$

Although written in Cartesian coordinates, quaternions can be written in any linearly−independent 4−basis because matrix algebra provides the necessary techniques for changing the basis. Therefore, like tensors, a quaternion equation is independent of the chosen basis. One could view quaternions as tensors restricted to a 4−dimensional algebraic field. For the sake of consistency, all transformations are also constrained to the same division algebra. This constraint might first appear too restrictive since for example it eliminates simple matrices for row permutations. Since quaternions are an algebraic field, there necessarily exists a combination of quaternions that achieves the action of a permutation. The need for consistency will overrule convenience.

Laws in physics are independent of coordinate systems. To make the coordinate independence explicit, amplitudes and basis vectors will be separated using a new notation. Consider a quaternion 4−function, A_n=(a_0, a_1, a_2, a_3), and an arbitrary 4−basis, Ihat_n=(ihat_0, ihat_1, ihat_2, ihat_3). In spacetime, the line that divides causality is define by the light cone. On the light cone, the total change in 3−space over the change in time is equal to one. Physics therefore indicates parity between the total 3−vector and the scalar, instead of weighing all four equally. A coordinate−independent Riemannian quaternion is defined to be A_0 Ihat_n=(a_0 ihat_0, a_1 ihat_1/3, a_2 ihat_2/3, a_3 ihat_3/3).

The equivalence principle of general relativity asserts, with experiments to back it up, that the inertial mass equals the gravitational mass. An accelerated reference frame can be indistinguishable from the effect of a mass density. No corresponding principle applies to electromagnetism, which depends only on the electromagnetic field tensor built from the potential. With Riemannian quaternions, the 4−unit vector does not have to be static, as illustrated by taking the time derivative of the first term and using the chain rule:

$$\frac{\partial a_0 \hat{i}_0}{\partial i_0} = \hat{i}_0 \frac{\partial a_0}{\partial i_0} + a_0 \frac{\partial \hat{i}_0}{\partial i_0}$$

The unit vector for time, ihat_0, can change over an infinitely small amount of time, i_0. Any change in a quaternion potential function could be due to contributions from a change in potential, the ihat_0 da_0/di_0 term, and/or a change in the basis, the a_0 dihat_0/di_0 term. Is this mathematical property related to physics? Consider Gauss' law written with Riemannian quaternions:

$$-\frac{\hat{i}_n^{\,2}}{9}\frac{\partial E_n}{\partial i_n}-\frac{\hat{i}_n\,E_n}{9}\frac{\partial \hat{i}_n}{\partial i_n}=4\pi\rho,\ \ n=1,2,3$$

The divergence of the electric field might equal the source, or equivalently, the divergence of the basis vectors. The "general equivalence principle" as defined here means that any measurement can be due to a change in the potential and/or a change in the basis vectors. The general equivalence principle is applicable to both gravity and electromagnetism.

■ Metrics and Quaternion Products

The theories of special and general relativity dictate the distance between events in spacetime. Although fundamentally different in their mathematical structure, inertia is a link between the two. Special relativity dictates the transformation rules for observers who change their inertia, assuming the system observed does not change. The field equations of general relativity detail the changes in distance due to a system changing its inertia from the vacuum to a non−zero energy density. A quaternion product necessarily contains information about the metric, but also has information in the 3−vector. This additional information about quaternion products will suggest a provocative link between metrics and inertia consistent with both special and general relativity.

Most structures in Nature do not transform like a scalar and a 3−vector. Quaternion products multiply two 4−basis vectors, and those products will transform differently. The rules of quaternion multiplication mirror those of complex numbers. Instead of the imaginary number i, there is a unit 3−vector for each quaternion playing an analogous role. The difference is that unit 3−vectors do not all have to point in the same direction. Based on the angle between them, two different unit 3−vectors have both a dot and cross product. The dot and cross products completely characterize the relationship between the two unit vectors. Compare the product of multiplying two complex numbers (a, bi) and (c, di):

$$(a,\ bi)\ (c,\ di) = (ac - bd,\ ad + bc),$$

with two quaternions, (a, B $\vec{I}$) and (c, D $\vec{I}$),

$$(a,\ B\,\hat{I})\ (c,\ D\,\hat{I}') = (ac - BD\,\hat{I}\cdot\hat{I}',\ aD\,\hat{I}' + Bc\,\hat{I} + BD\,\hat{I}\times\hat{I}')$$

Complex numbers commute because they do not have a cross product in the result. If the order of quaternion multiplication is reversed, then only the cross product would change its sign. Quaternion multiplication does not commute due to the behavior of the cross product. If the cross product is zero, then quaternion multiplication has all of the properties of complex numbers. If, on the other hand, the only value of a quaternion product is equal to the cross product, then multiplication is anti−commutative. Individually, the mathematical properties of commuting and anti−commuting algebras are well known. A quaternion product is the superposition of these two types of algebras that forms a division algebra.

Several steps are required to square of the difference of two Riemannian quaternions to form a measure of distance. First, the basis of the two quaternions must be shared. It makes no sense to subtract something in spherical coordinates from something in Cartesian coordinates. The basis does not have to be constant, only shared. Every quaternion commutes with itself, so the cross product is zero. There are seven unique pairs of basis vectors in a square:

$$\left(da_0\,\hat{i}_0,\ dA_n\,\frac{\hat{I}_n}{3}\right)^2 = \left(da_0^{\,2}\,\hat{i}_0^{\,2} - dA_n^{\,2}\,\frac{\hat{I}_n^{\,2}}{9},\ 2\,da_0\,dA_n\,\frac{\hat{i}_0\,\hat{I}_n}{3}\right)$$

The signs were chosen to be consistent with Hamilton's quaternion algebra. The four square basis vectors i_mu^2 define the metric. If the basis vectors are not constant, then the metric is dynamic. Define a "3−rope" to be the three other terms, which have the form i_0 I_n. Notice that the 3−rope starts in one time−space location and will have a non−zero length if it ends up at a different location and time. With quaternion products, the 3−rope is a natural companion to a metric for information about distance.

In special relativity, if the inertia of the observer but not the system is changed, the metric is invariant. The 3−rope is covariant, because it is known how it changes. A complementary hypothesis to the invariant metric of special relativity would propose the if the inertia of the system but not the observer is changed, there exists a choice of basis vectors such that the 3−rope is invariant but the metric changes in a known way. This could be written in algebraically using the following rule:

$$\hat{i}_0^{\,2} = \frac{-1}{\hat{I}_n^{\,2}}, \quad \left| \hat{i}_0 \hat{I}_n \right| = 1$$

If the magnitude of the time and 3−space basis vectors are inversely related, the magnitude of the product of the time basis vector with each 3−space basis vector will be constant even if the basis vectors themselves are dynamic. This hypothesis asserts there exists such a basis, but that particular basis does not have to be used.

Hamilton had the freedom to use the rule found in the above equation, but made the more obvious choice of i_0^2 = −I_n^2. The existence of a basis where the 3−rope is constant despite a change in the inertia of the system will have to be treated as provisional in this paper. In the second paper of this series, a metric with this property will be found and discussed.

■ Physically Relevant Differential Equations

Is there a rational way to construct physically relevant quaternion equations? The method used here will be to mimic the tensor equations of electromagnetism. The electromagnetic field strength tensor is formed by a differential operator acting on a potential. The Maxwell equations are formed by acting on the field with another differential operator. The Lorentz 4−force is created by the product of a electric charge, the electromagnetic field strength tensor, and a 4−velocity. This pattern will be repeated starting from an asymmetric field to create the same field and force equations using quaternion differentials and potentials. The challenge in this exercise is in the interpretation, to see how every term connects to established laws of physics.

As a first step to constructing differential equations, examine how the differential operator (d/dt, Del) acts on a potential function (phi, A):

$$\left(\frac{\partial}{\partial t}, \vec{\nabla} \right) (\phi, \vec{A}) = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla}.\vec{A}, \ \frac{\partial \vec{A}}{\partial t} + \vec{\nabla}\phi + \vec{\nabla} \times \vec{A} \right)$$

For the sake of clarity, the notation introduced for Riemann quaternions has been suppress, so the reader is encouraged to recognize that there are also a parallel set of terms for changes in the basis vectors. The previous equation is a complete assessment of the change in the 4−dimensional potential/basis, involving two time derivatives, the divergence, the gradient and the curl all in one. A unified field theory should account for all conceivable forms of change in a 4−dimensional potential/basis, as is the case here.

Quaternion operators and potentials have not been used to express the Maxwell equations. The reason can be found in the previous equation, where the sign of the divergence of A is opposite of the curl of A. In the Maxwell equations, the divergence and the curl involving the electric and magnetic field are all positive. Many others, even in Maxwell's time, have used complex−valued quaternions for the task because the extra imaginary number can be used to get the signs correct.

However, complex−valued quaternions are not an algebraic field. The norm, t^2+x^2+y^2+z^2, for a non−zero quaternion could equal zero if the values of t, x, y, and z were complex. This paper involves the constraint of working exclusively with 4−dimensional algebraic fields. Therefore, no matter how salutary the work with complex−valued quaternions, it is not relevant to this paper.

The reason to hope for unification using quaternions can be found in an analysis of symmetry provided by Albert Einstein:

"The physical world is represented as a four−dimensional continuum. If in this I adopt a Riemannian metric, and look for the simplest laws which such a metric can satisfy, I arrive at the relativistic gravitation theory of empty space. If I adopt in this space a vector field, or the antisymmetric tensor field derived from it, and if I look for the simplest laws which such a field can satisfy, I arrive at the Maxwell equations for free space." [einstein1934]

The "four−dimensional continuum" could be viewed as a technical constraint involving topology. Fortunately, quaternions do have a topological structure since they have a norm. Nature is asymmetric, containing both a symmetric metric for gravity and an antisymmetric tensor for electromagnetism. With this in mind, rewrite out the real 4x4 matrix representation of a quaternion:

$$q(t, x, y, z) = \begin{pmatrix} t & 0 & 0 & 0 \\ 0 & t & 0 & 0 \\ 0 & 0 & t & 0 \\ 0 & 0 & 0 & t \end{pmatrix} + \begin{pmatrix} 0 & -x & -y & -z \\ x & 0 & -z & y \\ y & z & 0 & -x \\ z & -y & x & 0 \end{pmatrix}$$

The scalar component (t in representation above) can be represented by a symmetric 4x4 matrix, invariant under transposition and conjugation (these are the same operations for quaternions). The 3−vector component (x, y and z in the representation above) is off−diagonal and can be represented by an antisymmetric 4x4 matrix, because taking the transpose will flip the signs of the 3−vector. Quaternions are asymmetric in their matrix representation, a property which is critical to using them for unifying gravity and electromagnetism.

■ Recreating the Maxwell Equations

Maxwell speculated that his set of equations might be expressed with quaternions someday (J. C. Maxwell, "Treatise on Electricity and Magnetism," Dover reprint, third edition, 1954). The divergence, gradient, and curl were initially developed by Hamilton during his investigation of quaternions. For the sake of logical consistency, any system of differential equations, such as the Maxwell equations, that depends on these tools must have a quaternion representation.

The Maxwell equations are gauge invariant. How can this property be built into a quaternion expression? Consider a common gauge such as the Lorenz gauge, dphi/dt + div A = 0. In quaternion parlance, this is a quaternion−scalar formed from a differential quaternion acting on a potential. To be invariant under an arbitrary gauge transformation, the quaternion−scalar must be set to zero. This can be done with the vector operator, (q−q*)/2. Search for a combination of quaternion operators and potentials that generate the Maxwell equations:

$$\frac{(\Box^* \text{Vector}(\Box^* A^*) - \Box \text{Vector}(\Box A))^*}{2} =$$

$$= \left(\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}),\ \frac{\partial}{\partial t}\left(\frac{\partial \vec{A}}{\partial t} + \vec{\nabla}\phi\right) + \vec{\nabla} \times (\vec{\nabla} \times \vec{A})\right) =$$

$$= \left(\vec{\nabla} \cdot \vec{B},\ -\frac{\partial \vec{E}}{\partial t} + \vec{\nabla} \times \vec{B}\right) =$$

$$= (0, 4\pi \vec{J}).$$

This is Ampere's law and the no monopoles vector identity (assuming a simply−connected topology). Any choice of gauge will not make a contribution due to the vector operator. If the vector operator was not used, then the gradient of the symmetric−matrix force field would be linked to the electromagnetic source equation, Ampere's law.

Generate the other two Maxwell equations:

$$\frac{-(\Box \text{Vector}(\Box^* A^*) + \Box^* \text{Vector}(\Box A))^*}{2} =$$

$$= \left(\vec{\nabla} \cdot \left(-\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi\right), \frac{\partial \vec{\nabla} \times \vec{A}}{\partial t} + \vec{\nabla} \times \left(-\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi\right)\right) =$$

$$= \left(\vec{\nabla} \cdot \vec{E}, \frac{\partial \vec{B}}{\partial t} + \vec{\nabla} \times \vec{E}\right) =$$

$$= \left(4\pi\rho, \vec{0}\right).$$

This is Gauss' and Faraday's law. Again, if the vector operator had not been used, the time derivative of the symmetric−matrix force field would be associated with the electromagnetic source equation, Gauss' law. To specify the Maxwell equations completely, two quaternion equations are required, just like the 4−vector approach.

Although successful, the quaternion expression is unappealing for reasons of simplicity, consistency and completeness. A complicated collection of sums or differences of differential operators acting on potentials − along with their conjugates − is required. There is no obvious reason this combination of terms should be central to the nature of light. One motivation for the search for a unified potential field involves simplifying the above expressions.

When a quaternion differential acts on a function, the divergence always has a sign opposite the curl. The opposite situation applies to the Maxwell equations. Of course the signs of the Maxwell equations cannot be changed. However, it may be worth the effort to explore equations with sign conventions consistent with the quaternion algebra, where the operators for divergence and curl were conceived.

Information about the change in the potential is explicitly discarded by the vector operator. Justification comes from the plea for gauge symmetry, essential for the Maxwell equations. The Maxwell equations apply to massless particles. Gauge symmetry is broken for massive fields. More information about the potential might be used in unification of electromagnetism with gravity. A gauge is also matrix symmetric, so it could provide a complete picture concerning symmetry.

■ One Unified Force Field from One Potential Field

For massless particles, the Maxwell equations are sufficient to explain classical and quantum electrodynamic phenomena in a gauge−invariant way. To unify electromagnetism with gravity, the gauge symmetry must be broken, opening the door to massive particles. Because of the constraints imposed by quaternion algebra, there is little freedom to choose the gauge with a simple quaternion expression. In the standard approach to the electromagnetic field, a differential 4−vector acts on a 4−vector potential in such a way as to create an antisymmetric second−rank tensor. The unified field hypothesis proposed involves a quaternion differential operator acting on a quaternion potential:

$$\square^* A^* = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla}.\vec{A}, \; - \frac{\partial \vec{A}}{\partial t} - \vec{\nabla} \phi + \vec{\nabla} \times \vec{A} \right)$$

This is a natural suggestion with this algebra. The antisymmetric−matrix component of the unified field has the same elements as the standard electromagnetic field tensor. Define the electric field E as the even terms, the ones that will not change signs if the order of the differential operator and the potential are reversed. The magnetic field B is the curl of A, the odd term. The justification for proposing the unified force field hypothesis rests on the presence of the electric and magnetic fields.

In some ways, the above equation looks just like the old idea of combining a scalar gauge field with the electromagnetic field strength tensor, as Gupta did in 1950 in order to quantize the Maxwell equations. He concluded that although useful because it is written in manifestly relativistic form, no new results beyond the Maxwell equation are obtained. Examine just the gauge contribution to the Lagrangian for this unified field:

$$L = -\frac{1}{2} \left(\frac{\partial \phi}{\partial t}, \; -\frac{1}{3} \frac{\partial A_x}{\partial x}, \; -\frac{1}{3} \frac{\partial A_y}{\partial y}, \; -\frac{1}{3} \frac{\partial A_z}{\partial z} \right)^2$$

Take the derivative of the Lagrangian with respect to the gauge variables:

$$\frac{\partial L}{\partial \frac{\partial A_\mu}{\partial x_\mu}} = 0$$

By Noether’s theorem, this conserved current indicates a symmetry of the Lagrangian. This is why the proposal involves new physics. The gauge is a dynamic variable constrained by the Lagrangian.

A quaternion potential function has four degrees of freedom represented by the scalar function phi and the 3−vector function A. Acting on this with one[or more] differential operators does not change the degrees of freedom. Instead, the tangent spaces of the potential will offer more subtle views on the rules for how potentials change.

The three classical force fields, g, E, and B, depend on the same quaternion potential, so there are only four degrees of freedom. With seven components to the three classical force fields, there must be three constraints between the fields. Two constraints are already familiar. The electric and magnetic field form a vector identity via Faraday’s law. Assuming spacetime is simply connected, the no monopoles equation is another identity. A new constraint arises because both the force fields for gravity and electricity are even. It will be shown subsequently how the even force fields can partially constructively or destructively interfere with each other.

■ Unified Field Equations

In the standard approach to generating the Maxwell equations, a differential operator acts on the electromagnetic field strength tensor. A unified field hypothesis for an isolated source is proposed which involves a differential quaternion operator acting on the unified field:

$$4 \pi (\rho, \vec{J})^* = \left(\frac{\partial}{\partial t}, \vec{\nabla} \right)^* \left(\frac{\partial}{\partial t}, \vec{\nabla} \right)^* (\phi, \vec{A})^* =$$

$$= \Bigg(\frac{\partial^2 \phi}{\partial t^2} - 2 \vec{\nabla} \cdot \frac{\partial \vec{A}}{\partial t} - \vec{\nabla} \cdot \vec{\nabla} \phi + \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}),$$

$$-2\vec{\nabla}\frac{\partial\phi}{\partial t} + \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) - \frac{\partial^2\vec{A}}{\partial t^2} - \vec{\nabla}\times(\vec{\nabla}\times\vec{A}) +$$

$$+2\vec{\nabla}\times\frac{\partial\vec{A}}{\partial t} + \vec{\nabla}\times\vec{\nabla}\phi\Big).$$

This second order set of four partial differential equations has four unknowns so this is a complete set of field equations. Rewrite the equations above in terms of the classical force fields:

$$4\pi(\rho,\ \vec{J})^* = \left(\frac{\partial g}{\partial t} + \vec{\nabla}.\vec{E} + \vec{\nabla}.\vec{B},\right.$$

$$\left.-\vec{\nabla}g + \frac{\partial\vec{E}}{\partial t} - \vec{\nabla}\times\vec{B} + \frac{\partial\vec{B}}{\partial t} - \vec{\nabla}\times\vec{E}\right).$$

The unified field equations contain three of the four Maxwell equations explicitly: Gauss' law, the no magnetic monopoles law, and Ampere's law. Faraday's law is a vector identity, so it is still true implicitly. Therefore, a subset of the unified field equations contains a quaternion representation of the Maxwell equations. The presence of the Maxwell equations justifies the investigation of the unified field equations.

There is a simple relationship between Faraday's law and the equation above. All that needs to be done is to subtract twice the time derivative of the magnetic field from both sides. What does this do to the 4−vector current density J? Now there is a current that transforms like a pseudo−current density. The inclusion of a pseudo−current along with a current making the proposal more complete. The volume integral of this pseudo−current density is the total magnetic flux:

$$k\iiint\frac{\partial B}{\partial t}\,dV = \frac{e}{\hbar c}\Phi_B$$

The unified field equation postulates a pseudo 3−vector current composed of the difference between the time derivative of the magnetic field and the curl of the electric field. The Aharonov−Bohm effect (Y Aharonov and D. Bohm, "Significance of electromagnetic potentials in the quantum theory," Phys. Rev, 115:485−491, 1959) depends on the total magnetic flux to create changes seen in the energy spectrum. The volume integral of the time derivative of the magnetic field is a measure of the total magnetic flux. The pseudo−current density is quite unusual, transforming differently under space inversion than the electric current density. One might imagine that a Lorentz transformation would shift this pseudo−current density into a pseudo−charge density. This does not happen however, because the vector identity involving the divergence of a curl still applies. The Aharonov−Bohm phenomenon, first viewed as a purely quantum effect, may have a classical analogue in the unified field equations.

The field equations involving the gravitational force field are dynamic and depend on four dimensions. This makes them likely to be consistent with special relativity. Since they are generated alongside the Maxwell equations, one can reasonably expect the differential equations will share many properties, with the ones involving the symmetric−matrix gravitational force field being more symmetric than those of the electromagnetic counterpart.

The unified source can be defined in terms of more familiar charge and current densities by separately setting the gravity or electromagnetic field equal to zero. In these cases, the source is due only to electricity or mass respectively. This leads to connections between the unified source, mass, and charge:

$$J = J_m \text{ iff } \vec{E} = \vec{B} = \vec{0}$$

$$J = J_e + \vec{J}_{AB} \quad \text{iff} \quad g = 0.$$

It would be incorrect − but almost true − to say that the unified charge and current are simply the sum of the three: mass, electric charge, and the Aharanov−Bohm pseudo−current (or total magnetic flux over the volume). These terms constructively interfere with each other, so they may not be viewed as being linearly independent.

Up to four linearly independent unified field equations can be formulated. A different set could be created by using the differential operator without taking its conjugate:

$$4\pi J^* = \square\square^* A^* =$$

$$= \left(\frac{\partial^2 \phi}{\partial t^2} + \vec{\nabla} \cdot \vec{\nabla}\phi - \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}),\ -\frac{\partial^2 \vec{A}}{\partial t^2} - \vec{\nabla}^2 \vec{A} - \vec{\nabla} \times \vec{\nabla}\phi \right)$$

$$= \left(\frac{\partial g}{\partial t} - \vec{\nabla}.\vec{E} - \vec{\nabla}.\vec{B}, \right.$$

$$\left. \vec{\nabla} g + \frac{\partial \vec{E}}{\partial t} + \vec{\nabla} \times \vec{B} + \frac{\partial \vec{B}}{\partial t} + \vec{\nabla} \times \vec{E} \right).$$

This is an elliptic equation. Since the goal of this work is a complete system of field equations, this may turn out to be an advantage. An elliptic equation combined with a hyperbolic one might more fully describe gravitational and electromagnetic waves from sources. Unlike the first set of field equations, the cross terms destructively interfere with each other.

The elliptic field equation again contains three of four Maxwell equations explicitly:Gauss' law, the no magnetic monopoles vector identity and Faraday's law. This time, Ampere's law looks different. To be consistent with Ampere's law, again a pseudo−current must be included. This may be the differential form of a classical Aharonov−Bohm effect.

The only term that does not change between the two field equations is the one involving the dynamic gravitational force. This might be a clue for why this force is only attractive.

■ Solutions to the Unified Field Equations

All the solutions that have been worked out for the Maxwell equations will work with the unified field equations. For example, if the potential is static, the scalar equation for hyperbolic field equation is the Poisson equation. The unified equations are more informative, since any potential which is a solution to the scalar Poisson equation will also characterize the corresponding current.

The field equations of general relativity and the Maxwell equations both have vacuum solutions, such as plane wave solutions. The unified field equations do not have such a solution, other than a constant. Given historical tradition, this may seem like a deadly flaw. However, it may be something that is required for a final and complete theory. In a unified field theory, the gravitational part may be zero while the electrical part is not, and visa versa. Non−zero solutions are worth exploring

An inverse square potential plays an important role in both gravity and electromagnetism. Examine the scalar field involving the inverse interval squared:

$$\square\square^* \left(\frac{1}{t^2 - x^2 - y^2 - z^2},\ \vec{0} \right) = \left(\frac{4\,(3\,t^2 + x^2 + y^2 + z^2)}{(t^2 - x^2 - y^2 - z^2)^3},\ \vec{0} \right)$$

This potential solves the Maxwell equations in the Lorentz gauge:

$$\Box^2 \left(\frac{1}{t^2 - x^2 - y^2 - z^2}, \vec{0} \right) = 0$$

The non−zero part may have everything to do with gravity.

A plane wave solution does exist, but not for a pure vacuum. Instead, a plane wave solution exists with the constrain that the net current is zero for the elliptical field equations

The field equations of general relativity and the Maxwell equations both have vacuum solutions. A vacuum solution for the unified field equation is apparent for the elliptical field equations:

$$A = \left(\phi_0 \, e^{\vec{K}\cdot\vec{R}-\omega t}, \vec{A}_0 \, e^{\vec{K}\cdot\vec{R}-\omega t} \right)$$

The unified field equation will evaluate to zero if

$$\text{Scalar}\left(\left(\frac{\omega}{c}, \vec{K} \right)^2 \right) = 0$$

The dispersion relation is an inverted distance, so it will depend on the metric. The same potential can also solve the hyperbolic field equations under different constraints and resulting dispersion equation (not shown). There were two reasons for not including the customary imaginary number "i" in the exponential of the potential. First, it was not necessary. Second, it would have created a complex−valued quaternion, and therefore is outside the domain of my work. The important thing to realize is that vacuum solutions to the unified field equations exist whose dispersion equations depend on the metric. This is an indication that unifying gravity and electromagnetism is an appropriate goal.

■ Conservation Laws

Conservation of electric charge is implicit in the Maxwell equations. Is there also a conserved quantity for the gravitational field? Examine how the differential operator acts on the unified field equation:

$$\Box\,\Box^*\,\Box^*\,A^* = \left(\frac{\partial^2 g}{\partial t^2} + \vec{\nabla} \cdot \vec{\nabla} g, \; \frac{\partial^2 \vec{E}}{\partial t^2} + \vec{\nabla}^2 \vec{E} + \frac{\partial^2 \vec{B}}{\partial t^2} + \vec{\nabla}^2 \vec{B} \right)$$

Notice that the gravitational force field only appears in the quaternion scalar. The electromagnetic fields only appear in the 3−vector. This generates two types of constraints on the sources. No change in the electric source applies to the quaternion scalar. No change in the gravitational source applies to the 3−vector.

$$\text{Scalar}(\Box J_e{}^*) = \frac{\partial \rho_e}{\partial t} + \vec{\nabla} \cdot \vec{J}_e = 0$$

$$\text{Scalar}\left(\Box \vec{J}_{AB}{}^*\right) = \nabla \cdot \vec{J}_{AB}{}^* = 0$$

$$\text{Vector}(\Box J_m{}^*) = -\frac{\partial \vec{J}_m}{\partial t} + \vec{\nabla} \rho_m - \vec{\nabla} \times \vec{J}_m = \vec{0}$$

The first equation is known as the continuity equation, and is the reason that electric charge is conserved. For a different inertial observer, this will appear as a conservation of electric current density. There is no source term for the Aharanov−Bohm current, and subsequently no conservation law. The 3−vector equation is a constraint on the mass current density, and is the reason mass current density is conserved. For a different inertial observer, the mass density is conserved.

■ Transformations of the Unified Force Field

The transformation properties of the unified field promise to be more intricate than either gravity or electromagnetism separately. What might be expected to happen under a Lorentz transformation? Gravity involves mass that is Lorentz invariant, so the field that generates it should be Lorentz invariant. The electromagnetic field is Lorentz covariant. However, a transformation cannot do both perfectly. The reason is that a Lorentz transformation mixes a quaternion scalar with a 3−vector. If a transformation left the quaternion scalar invariant and the 3−vector covariant, the two would effectively not mix. The effect of unification must be subtle, since the transformation properties are well known experimentally.

Consider a boost along the x−axis. The gravitational force field is Lorentz invariant. All the terms required to make the electromagnetic field covariant under a Lorentz transformation are present, but covariance of the electromagnetic fields requires the following residual terms:

$$(\square'^{*} A'^{*})_{\text{Residual}} = \Big(0,\ (\gamma^2 \beta^2 - 1)\ \frac{\partial A_x}{\partial t} + (\gamma^2 - 1)\ \frac{\partial \phi}{\partial x},$$

$$-2\,\gamma\beta\left(\frac{\partial A_z}{\partial t} + \frac{\partial A_y}{\partial x}\right),\ 2\,\gamma\beta\left(\frac{\partial A_y}{\partial t} - \frac{\partial A_z}{\partial x}\right)\Big).$$

At this time, the correct interpretation of the residual term is unclear. Most importantly, it was shown earlier that charge is conserved. These terms could be a velocity−dependent phase factor. If so, it might provide a test for the theory.

The mechanics of the Lorentz transformation itself might require careful re−examination when so strictly confined to quaternion algebra. For a boost along the x−axis, if only the differential transformation is in the opposite direction, then the electromagnetic field is Lorentz covariant with the residual term residing with the gravitational field. The meaning of this observation is even less clear. Only relatively recently has DeLeo been able to represent the Lorentz group using real quaternions (S. De Leo, "Quaternions and special relativity," J. Math. Phys., 37(6):2955−2968, 1996). The delay appears odd since the interval of special relativity is the scalar of the square of the difference between two events. In the real 4x4 matrix representation, the interval is a quarter of the trace of the square. Therefore, any matrix with a trace of one that does not distort the length of the scalar and 3−vector can multiply a quaternion without effecting the interval. One such class is 3−dimensional, spatial rotations. An operator that adds nothing to the trace but distorts the lengths of the scalar and 3−vector with the constraint that the difference in lengths is constant will also suffice. These are boosts in an inertial reference frame. Boosts plus rotations form the Lorentz group.

Three types of gauge transformations will be investigated: a scalar, a 3−vector, and a quaternion gauge field. Consider an arbitrary scalar field transformation of the potential:

$$A \to A' = A - \square^{*} \lambda.$$

The electromagnetic fields are invariant under this transformation. An additional constraint on the gauge field is required to leave the gravitational force field invariant, namely that the scalar gauge field solves a homogeneous elliptical equation. From the perspective of this proposal, the freedom to choose a scalar gauge field for the Maxwell equations is due to the omission of the gravitational force field.

Transform the potential with an arbitrary 3−vector field:

$$A \to A' = A - \Box^* \vec{\Lambda}.$$

This time the gravitational force field is invariant under a 3−vector gauge field transformation. Additional constraints can be placed on the 3−vector gauge field to preserve a chosen electromagnetic invariant. For example, if the difference between the two electromagnetic fields is to remain invariant, then the 3−vector gauge field must be the solution to an elliptical equation. Other classes of invariants could be examined.

The scalar and 3−vector gauge fields could be combined to form a quaternion gauge field. This gauge transformation would have the same constraints as those above to leave the fields invariant. Is there any such gauge field? The quaternion gauge field can be represented the following way:

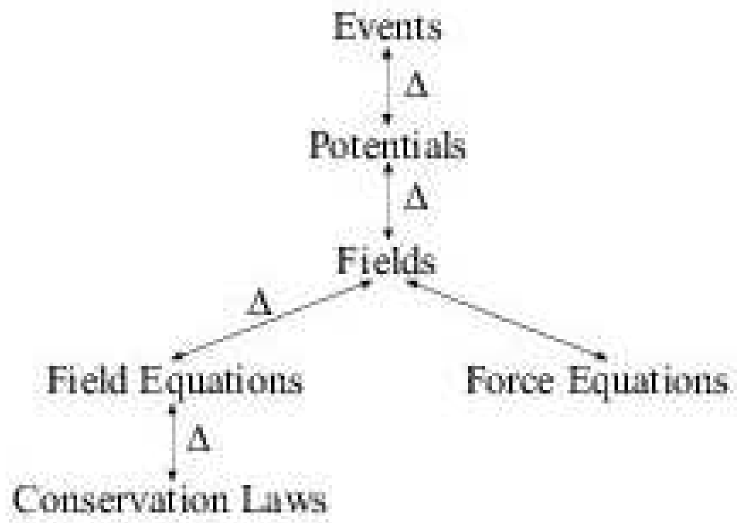
$$A \to A' = A - \Box^* \Lambda.$$

If a force field is created by hitting this gauge transformation with a differential operator, then the gauge field becomes a unified field equation. Since vacuum solutions have been found for those equations, a quaternion gauge transformation can leave the field invariant.

■ Future Directions

The fields of gravity and electromagnetism were unified in a way consistent with Einstein's vision, not his technique. The guiding principles were simple but unusual: generate expressions familiar from electromagnetism using quaternions, striving to interpret any extra terms as being due to gravity. The first hypothesis about the unified field involved only a quaternion differential operator acting on a potential, no extra terms added by hand. It contained the typical potential representation of the electromagnetic field, along with a symmetric−matrix force field for gravity. The second hypothesis concerned a unified field equation formed by acting on the unified field with one more differential operator. All the Maxwell equations are included explicitly or implicitly. Additional terms suggested the inclusion of a classical representation of the Aharanov−Bohm effect. Four linearly independent unified field equations exist, but only the hyperbolic and elliptic cases were discussed. A large family of vacuum solutions exists, and will require future analysis to appreciate.

Why did this approach work? The hypothesis that initiated this line of research was that all events in spacetime could be represented by quaternions, no matter how the events were generated. This is a broad hypothesis, attempting to reach all areas covered by physics. Based on the equations presented in this paper, a logical structure can be constructed, starting from events (see figure below). A set of events forms a pattern that can be described by a potential. The change in a potential creates a field. The change in field creates a field equation. The terms that do not change under differentiation of a field equation form conservation laws.



■ Acknowledgements

Thanks for many productive discussions Prof. Guido Sandri.

Einstein's Vision II: A Unified Force Equation with Constant Velocity Solutions

■ Abstract

In quantum electrodynamics, photons have four modes of transmission, at least mathematically: two transverse modes for electrodynamics, a longitudinal, and a scalar mode. The probabilities of the last two modes cancel each other out for a spin 1 field, but that does not have to be the case for a field with spin 1 and spin 2 particles. One potential solution to the field equations is found which depend on the inverse of an interval between two events squared. The force field created by the potential is constructed by comparison with the classical Newtonian gravitational field. The Lagrange density L= scalar(−(J A*) − 1/2 Box* A Box A*) can contribute to a scalar mode, but still has the field equations of Maxwell with the choice of the Lorenz gauge. A relativistic force equation is proposed, created by the product of charge, normalized force field, and 4−velocity: dmU/dtau = kq Box* A/|A| U*. The solution to the force equation using the inverse square interval potential is found. Eliminating the constants generates a metric equation, $d\tau^2 = e^{-2\frac{GM}{c^2\tau}} - e^{2\frac{GM}{c^2\tau}} \, dR^2/c^2$, where tau is a lightlike interval with almost the same magnitude as the radius R of separation between source and test masses. For a weak gravitational field, the metric will pass the same tests as the Schwarzschild metric of general relativity. The two metrics differ for higher order terms, which makes the proposed metric distinct and testable experimentally. A constant−velocity solution exists for the gravitational force equation for a system with an exponentially−decaying mass distribution. The dark matter hypothesis is not needed to explain the constant−velocity profiles seen for some galaxies. Gravity is a metric theory, electromagnetism is not. By using Riemannian quaternions which can have dynamic basis vectors, it becomes possible to merge metric theory with the linear Maxwell equations. The proposal may also have implications for classical big bang theory.

■ An Opportunity for Classical Gravity?

The electrodynamic field can be quantized in a manifestly covariant form by fixing the gauge (K. Bleuler, Helv. Phys. Acta, 23:567, 1950, and S. N. Gupta, "Theory of longitudinal photons in quantum electrodynamics", Proc. Phys. Soc., 63:681−691). The starting point is the 4−potential A^mu. There are four modes of transmission for photons corresponding to the four degrees of freedom: two transverse, one scalar, and one longitudinal. Gupta calculated that "the probability of the emission of a real longitudinal photon is canceled by the 'negative probability' of the emission of a corresponding scalar photon." He notes that this does not always have to be the case for the nonhomogeneous Maxwell equations, which is the focus of this work. A scalar photon would not change signs under a space or time reversal, so its symmetry is different from the electric 3−vector field and the magnetic 3−pseudo−vector field, and thus does not have an obvious role to play in electrodynamics.

My hypothesis is that the scalar and longitudinal photons for the electromagnetic field constitute gravity. The hypothesis makes several predictions even at this preliminary stage. First, the math of gravity and electromagnetism should be similar but not identical. The inverse square form of Newton's law of gravity was a direct inspiration for Coulomb's law. Gravity should be more symmetric than electromagnetism because the mode is scalar, instead of transverse. The second rank field strength tensor in general relativity is symmetric while the analogous tensor for the electromagnetic field is antisymmetric. Since the mode of gravity is orthogonal to electromagnetism, the charges can be likewise, so there will be no simple relationship between gravitational charge (mass) and electric charge. Gravitational waves in general relativity are transverse, so this proposal is distinct from general relativity. Nature exploits all the math available, so it is unreasonable to suppose that scalar and longitudinal photons are never used for anything. Whatever phenomenon exploits the scalar and longitudinal photons must be similar, but just as important as electromagnetism. Gravity is a natural candidate.

■ A Gravitational Field Inside Maxwell

Newton's classical gravitational law arises from a scalar potential. Here is the scalar field equation:

$$\nabla^2 \phi = 4 \pi G \rho$$

For the case of a vacuum, when rho = 0, this is known as the Laplace equation. For a spherically symmetric source, one solution is:

$$\phi = -\frac{GM}{\sqrt{x^2 + y^2 + z^2}}$$

The problem with the field equation is that the Laplace operator does not have a time differential operator. Any change in the mass density propagates at infinite speed, in conflict with special relativity (MTW, chapter 7). One way to derive the field equations of general relativity involves making Newton's law of gravity consistent with the finite speed of light.

A way to repair the field equations is to use the D'Alembertian operator, which is four dimensional. That expression is identical to the A^0 component of the Maxwell equations with the choice of the Lorenz gauge. The sources are of course different. Yet the argument being made here is that there are degrees of freedom which have yet to be exploited. For the two degrees of freedom, we can have a different source term, mass:

$$\square^2 A = 4 \pi (J_q - J_m)$$

If one is studying scalar or longitudinal modes, the source is J_mass, the mass current density. If one is working with transverse modes, the source is J_charge^mu , the electric charge density. Since the modes are orthogonal, the sources can be also.

To be consistent with the classic scalar potential yet still be relativistic, the potential must have x^2, y^2, z^2, and t^2. This suggests a particular solution to the field equations:

$$A = \left(\frac{1}{c^2 t^2 - x^2 - y^2 - z^2}, \vec{0}\right) = \left(\frac{1}{\tau^2}, \vec{0}\right)$$

This potential is interesting for several reasons. It is the inverse of the Lorentz−invariant interval squared. Like mass, the 4−potential will not be altered by a change in an inertial reference frame. The interval between any two events will contribute to the potential. General relativity applies to any form of energy, including gravitational field energy. A potential that embraces every interval may have a broad enough scope to do the work of gravity.

The potential also has serious problems. Classical gravity depends on an inverse square force field, not an inverse square potential. Taking the derivative of the potential puts a forth power of the interval in the denominator. At this point, I could stop and say that this potential has nothing to do with gravity because it has the wrong dependence on distance. An alternative is to look for an algebraic way to repair the problem. This is the type of approach used by the early workers in quantum mechanics like de Broglie, and will be adopted here. The equations of motion can be normalized to the magnitude of the 4−potential:

$$\frac{\Box^2 A}{|A|} = 4\pi\,(k J_{charge} + G J_{mass})$$

Since the magnitude of the potential is the inverse interval squared, the resulting equation has only an interval squared in the denominator. An interval is not necessarily the same as the distance R between the source and test mass used in the classical theory. However, I can impose a selection rule that in the classical limit, the only events that contribute to the potential are those that are timelike separated between the source and the test masses. It takes a timelike interval to know that the source is a distance R away. Action−at−a−distance respects the speed of light as it must.

■ Search for the Source Mass

Where is the source mass in the potential? All that has been discussed so far is an interval, a distance, nothing about mass. An idea from general relativity will be borrowed, that mass can be treated geometrically if multiplied by the constants G/c^2. The distance between the Earth and the Sun is approximately 1.5x10^11 m, while the Sun's mass expressed in units of distance, GM_-Sun/c^2, is 1.5x10^3 m, eight orders of magnitude smaller. The overall length of the interval will not be changed noticeably if the spatial separation and the Sun's mass expressed as a distance are summed. However, the force field is the derivative of the potential, and any change in position in spacetime will have a far greater effect proportionally on the smaller geometric mass than the spatial separation. Make the following change of variables:

$$t \to t' = A + \frac{GM}{2c^2 A}\,t$$

$$\vec{R} \to \vec{R}' = \vec{B} + \frac{GM}{2c^2\,|\vec{B}|}\,\vec{R}$$

where A and B are locally constants such that tau^2 ~= A^2 − B^2. The change of variables is valid locally, but not globally, since it breaks down for arbitrarily long time or distance away. General relativity is also valid locally and not globally.

What is the physical interpretation of the inverse square potential and the above substitution? Newton observed that motion in an ellipse could be caused by either a linear central force or an inverse square law. With the above substitution, there is a linear displacement equation inside an inverse square potential. It is like a simple harmonic oscillator inside a simple harmonic oscillator! This oscillator works with four dimensions. Although it is confusing to confront the idea of oscillations in time, there is no need worry about it, since the equations are quite simple and their mathematical consequences can be worked out. If all the terms where included, the equation would be nonlinear.

The field is the derivative of the potential. To be correct technically, it is the contravariant derivative. This requires both a metric and a connection. In effect, all the work presented with quaternions uses the Minkowski metric with Cartesian coordinates. For such a choice of metric and coordinates, the contravariant derivative equals the normal derivative. The derivative of the potential under study, a normalized interval squared with the linear displacement substitution, is approximately:

$$\frac{1}{\left|\frac{1}{\tau^2}\right|}\,\frac{\partial \frac{1}{\tau^2}}{\partial t} = -\frac{GM}{c^2\,\tau^2}$$

$$\frac{1}{\left|\frac{1}{\tau^2}\right|}\frac{\partial\frac{1}{\tau^2}}{\partial\vec{R}} = \frac{GM}{c^2\,\tau^2}$$

This should look familiar, remembering that employing the event selection rule from above, the magnitude of tau ^2 is almost the same as R^2, differing only by the geometric mass of the source.

■ A Lagrangian for Four Modes

Despite its formulation using quaternions, this unification proposal is strikingly similar to earlier work. Gupta wanted to quantize the radiation field using a form that was manifestly covariant in its explicit treatment of time and space. He fixed the gauge with this Lagrange density:

$$L = -J^\mu A_\mu - \frac{1}{2}(\partial^\mu A_\mu)^2 - \frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu)$$

The equations of motion for this Lagrangian are the same as choosing the Lorenz gauge:

$$\Box^2 A^\mu = J^\mu$$

The problem with the Lagrangian is that the field strength tensor is antisymmetric. Due to the zeros along the diagonal, it cannot contribute directly to a scalar mode. What is needed is a Lagrange density that could contribute directly to the scalar mode but still have the same field equations. Here is such a Lagrangian:

$$L = \text{scalar}\left(-(J_q - J_m)A^* - \frac{1}{2}\Box^* A\,\Box A^*\right)$$

This is not as miraculous as it might first appear. It is the first of four terms generated in the contraction of the electromagnetic field strength tensor. In essence, information is not discarded, which is what happens in making the field strength tensor antisymmetric. The one remaining modification is to normalize both the Lagrangian and equations of motion to the size of the potential.

■ From a Relativistic 4−force to a Metric

A relativistic 4−force is the change in momentum with respect to the interval. The covariant force law is similar in form to the one for electromagnetism except that the second rank tensor is asymmetric and normalized:

$$F = \frac{\partial p}{\partial\tau} = mc\frac{\partial\beta}{\partial\tau} + \beta c\frac{\partial m}{\partial\tau} = k\,q\,\frac{\Box^* A^*}{|A|}\beta^*$$

If this equation is to transform like the Lorentz 4−force of electromagnetism, the normalized potential must be invariant under a Lorentz transformation. That is the case of the potential under study.

In the first application of the force law, assume the derivative of the mass with respect to the interval is zero. For the scalar photons, assume the charge q is the gravitational test mass. Experiments have demonstrated that gravitational and inertial masses are equal. Assuming spherical symmetry, the inverse interval squared potential leads to the following equations of motion:

$$\left(\frac{\partial^2 t}{\partial\tau^2} + \frac{GM}{c^2\,\tau^2}\frac{\partial t}{\partial\tau},\ \frac{\partial^2\vec{R}}{\partial\tau^2} - \frac{GM}{c^2\,\tau^2}\frac{\partial\vec{R}}{\partial\tau}\right) = \left(0,\ \vec{0}\right)$$

Solve these second−order differential equations for the spacetime position:

$$t = c_1 \left(\tau e^{\frac{GM}{c^2 \tau}} - \frac{GM}{c^2} \mathrm{Ei} \left(\frac{GM}{c^2 \tau} \right) \right) + c_2$$

$$\vec{R} = \vec{C}_1 \left(\tau e^{-\frac{GM}{c^2 \tau}} + \frac{GM}{c^2} \mathrm{Ei} \left(-\frac{GM}{c^2 \tau} \right) \right) + \vec{C}_2$$

where Ei is the exponential integral, Ei(t)=the integral from negative infinity to t of e^t/t dt. The exponential integral plays other roles in quantum mechanics, so its presence is interesting.

Eight constants need to be eliminated:(c_1, C_1) and (c_2, C_2). Take the derivative of the space-time position with respect to tau. This eliminates four constants, (c_2, C_2). The result is a 4−velocity:

$$\frac{\partial t}{\partial \tau} = c_1 \, e^{\frac{GM}{c^2 \tau}}$$

$$\frac{\partial \vec{R}}{\partial \tau} = \vec{C}_1 \, e^{-\frac{GM}{c^2 \tau}}$$

In flat spacetime, beta _mu beta^mu=1, providing four more constraints. Spacetime is flat if M goes to 0 or tau goes to infinity, leading to e^(GM/c^2|tau |) goes to 1:

$$\left(\frac{\partial t}{\partial \tau} \right)^2 - \left(\frac{\partial \vec{R}}{\partial \tau} \right) \cdot \left(\frac{\partial \vec{R}}{\partial \tau} \right) = c_1{}^2 - \vec{C}_1 \cdot \vec{C}_1 = 1$$

Solve for c_1^2 and C_1.C_1:

$$c_1{}^2 = e^{-\frac{GM}{c^2 \tau}} \; \frac{\partial t}{\partial \tau}$$

$$\vec{C}_1 \cdot \vec{C}_1 = e^{\frac{GM}{c^2 \tau}} \; \frac{\partial \vec{R}}{\partial \tau}$$

Substitute back into the flat spacetime constraint. Rearrange into a metric:

$$\partial \tau^2 = e^{-2 \frac{GM}{c^3 \tau}} \; \partial t^2 - e^{2 \frac{GM}{c^3 \tau}} \; \partial \vec{R}^2$$

If the gravitational field is zero, this generates the Minkowski metric of flat spacetime. Conversely, if the gravitational field is non−zero, spacetime is curved

As expected, this become the Minkowski metric for flat spacetime if M goes to 0 or tau goes to infinity.

No formal connection between this proposal and curvature has been established. Instead a path between a proposed gravitational force equation and a metric function was sketched. There is a historical precedence for the line of logic followed. Sir Isaac Newton in the Principia showed an important link between forces linear in position and inverse square force laws. More modern efforts have shown that the reason for the connection is due to the conformal mapping of z goes to z^2 (T. Needham, "Newton and the transmutation of force," Amer. Math. Mon., 100:119−137, 1993). This method was adapted to a quaternion force law linear in the relativistic velocity to generate a metric.

For a weak field, write the Taylor series expansion in terms of the total mass over the interval to second−order in M/|tau|:

$$\partial\tau^2 = \left(1 - 2\,\frac{G\,M}{c^2\,\tau} + 2\left(\frac{G\,M}{c^2\,\tau}\right)^2\right)\partial t^2 -$$

$$-\left(1 + 2\,\frac{G\,M}{c^2\,\tau} + 2\left(\frac{G\,M}{c^2\,\tau}\right)^2\right)\partial \vec{R}^2 + O\left(\left(\frac{G\,M}{c^2\,\tau}\right)^3\right)$$

Contrast this with the Schwarzschild solution in isotropic coordinates expanded to second order in M/R (MTW, eq. 31.22):

$$\partial\tau^2 = \left(1 - 2\,\frac{G\,M}{c^2\,R} + 2\left(\frac{G\,M}{c^2\,R}\right)^2\right)\partial t^2 -$$

$$-\left(1 + 2\,\frac{G\,M}{c^2\,R} + 2.5\left(\frac{G\,M}{c^2\,R}\right)^2\right)\partial \vec{R}^2 + O\left(\left(\frac{G\,M}{c^2\,R}\right)^3\right)$$

The magnitude of the lightlike interval tau in the unified field metric is nearly identical to the radius R in the Schwarzschild metric, the difference being the geometric mass of the source included in the interval tau . The metric for the scalar potential will pass the same weak field tests of general relativity as the Schwarzschild metric to post−Newtonian accuracy, which does not use the second order spatial term. The difference in the higher order terms can be the basis of an experimental test to distinguish this proposal from general relativity. Since the effect is second order in the field strength, such a test will challenge experimental techniques.

The two metrics are numerically very similar for weak fields, but mathematically distinct. For example, the Schwarzschild metric is static, but the unified metric contains a dependence on time so is dynamic. The Schwarzschild metric has a singularity at R=0. The unified gravitational force metric becomes undefined for lightlike intervals. This might pose less of a conceptual problem, since light has no rest mass.

■ The Constant Velocity Profile Solution

In the previous section, the system had a constant point−source mass with a velocity profile that decayed with distance. Here the opposite situation is examined, where the velocity profile is a constant, but the mass distribution decays with distance. Expand the definition of the relativistic force using the chain rule:

$$c\,\frac{\partial m\,\beta}{\partial\tau} = m\,c\,\frac{\partial\beta}{\partial\tau} + \beta\,c\,\frac{\partial m}{\partial\tau}$$

The first term of the force is the one that leads to an approximation of the Schwarzschild metric, and by extension, Newton’s law of gravity. For a region of spacetime where the velocity is constant, this term is zero. In that region, gravity’s effect is on the distribution of mass over spacetime. This new gravitational term is not due to the unified field proposal per se. It is more in keeping with the principles underlying relativity, looking for changes in all components, in this case mass distribution with respect to spacetime.

Start with the gravitational force in a region of spacetime with no velocity change:

$$\beta\,c\,\frac{\partial m_i}{\partial\tau} = k\,m_g\,\mathrm{Scalar}\left(\Box^*\,A^*\right)\,\beta^*$$

Make the same assumptions as before: the gravitational mass is equal to the inertial mass and the gravitational field employs the interval between the worldlines of the test and gravitational masses. This generates an equation for the distribution of mass:

$$\left(\gamma \frac{\partial m}{\partial \tau} + \frac{\gamma G M}{c^2 \mid \tau \mid^2} m, \gamma \vec{\beta} \frac{\partial m}{\partial \tau} - \frac{\gamma \vec{\beta} G M}{c^2 \mid \tau \mid^2} m\right) = \left(0, \vec{0}\right)$$

Solve for the mass flow:

$$\left(\gamma m, \gamma \vec{\beta} m\right) = \left(c\, e^{\frac{G M}{c^2 |\tau|}}, \vec{C}\, e^{-\frac{G M}{c^2 |\tau|}}\right)$$

As in the previous example for a classical weak field, assume the magnitude of the interval is an excellent approximation to the radius divided by the speed of light. The velocity is a constant, so it is the mass distribution that shows an exponential decay with respect to the interval, which is numerically no different from the radius over the speed of light. This is a stable solution. If the mass keeps dropping of exponentially, the velocity profile will remain constant

Look at the problem in reverse. The distribution of matter has an exponential decay with distance from the center. It must solve a differential equation with the velocity constant over that region of spacetime like the one proposed.

The exponential decay of the mass of a disk galaxy is only one solution to this expanded gravitational force equation. The behavior of larger systems, such as gravitational lensing caused by clusters, cannot be explained by the Newtonian term (A. G. Bergmann, V. Petrosian, and R. Lynds, "Gravitational lens images of arcs in clusters," Astrophys. J., 350:23, 1990. S. A. Grossman and R. Narayan, "Gravitationally lensed images in abell 370," Astrophys. J., 344−637−644, 1989. J. A. Tyson, F. Valdes, and R. A. Wenk, "Detection of systematic gravitational lens galaxy image alignments: Mapping dark matter in galaxy clusters," Astrophys. J. Let., 349:L1, 1990). It will remain to be seen if this proposal is sufficient to work on that scale.

■ Metrics and Forces

Gravity was first described as a force by Isaac Newton. In general relativity, Albert Einstein argued that gravity was not a force at all. Rather, gravity was Riemannian geometry, curvature of spacetime caused by the presence of a mass−energy density. Electromagnetism was first described as a force, modeled on gravity. That remains a valid choice today. However, electromagnetism cannot be depicted in purely geometric terms. A conceptual gap exists between purely geometrical and force laws.

The general equivalence principle, introduced in the first paper of this series, places geometry and force potentials on equal footing. Riemannian quaternions, (a_0 i_0, a_1 i_1/3, a_2 i_2/3, a_3 i_3/3), has pairs of (possibly) dynamic terms for the 4−potential A and the 4−basis I. Gauss' law written with Riemannian quaternion potentials and operators leads to this expression:

$$-\frac{\hat{i}_n^{\,2}}{9} \frac{\partial E_n}{\partial i_n} - \frac{\hat{i}_n E_n}{9} \frac{\partial \hat{i}_n}{\partial i_n} = 4 \pi \rho, \quad n = 1, 2, 3$$

If the divergence of the electric field E was zero, then Gauss' law would be due entirely to the divergence of the basis vectors. The reverse case could also hold. Any law of electrodynamics written with Riemannian quaternions is a combination of changes in potentials and/or basis vectors.

■ Future Directions

An algebraic path between a solution to the Maxwell equations and a classical metric gravitational theory has been shown. No effort has been extended yet to quantize the unification proposal. Like the early work in quantum mechanics, a collection of hunches is used to connect equations. One is left with the question of why this might work? The action of a gauge invariant theory cannot be inverted to generate the propagator needed for quantum mechanics. Fixing the gauge makes the action invertible, but the additional constraint decreases the degrees of freedom. By using quaternions, a division algebra, the equation is necessarily invertible without imposing a constraint. If the operation of multiplication surpasses what can be done with division, then Nature cannot harness the most robust mathematical structure, a topological algebraic field, the foundation for doing calculus. Nature does calculus in four dimensions, and it is this requirement that fixes the gauge. In the future, when we understand how to do calculus with four−dimensional automorphic functions, we may have a deep appreciation of Nature’s methods.

There is a physical explanation for gravity − it is a local, nonlinear, four−dimensional simple harmonic oscillator. Gravity is all about oscillations. The Earth returns to approximately the same place after one year of travel. If there were no interfering matter in the way, an apple dropped would fall to the center of the Earth, reach the other side, and return in a little over eighty minutes. The metric equation that results from this analysis is within the experimental constraints of current tests of general relativity. That makes the proposal reasonable. For higher order terms of a weak field, the proposal is different than the Schwarzschild metric of general relativity. That makes it testable. There are very few reasonable, testable classical unified field theories in physics, so this alone should spark interest in this line of work.

For a spiral galaxy with an exponential mass distribution, dark matter is no longer needed to explain the flat velocity profile observed or the long term stability of the disk. Mass distributed over large distances of space has an effect on the mass distribution itself. This raises an interesting question: is there also an effect of mass distributed over large amounts of time? If the answer is yes, then this might solve two analogous riddles involving large time scales, flat velocity profiles and the stability of solutions. Classical big bang cosmology theory spans the largest time frame possible and faces two such issues. The horizon problem involves the extremely consistent velocity profile across parts of the Universe that are not casually linked (MTW, p. 815). The flatness problem indicates how unstable the classical big bang theory is, requiring exceptional fine tuning to avoid collapse. Considerable effort will be required to substantiate this tenuous hypothesis. Any insight into the origin of the unified engine driving the Universe of gravity and light is worthwhile.

Strings and Quantum Gravity

In this section, a quaternion 3−string will be defined. By making this quantity dimensionless, I will argue that it my be involved in a relativistic quantum gravity theory, at least one consistent with current experimental tests. At the current time, this is an idea in progress, not a theory, since the equations of motion have not been determined. It is hoped that the work in the previous section on unified fields will provide that someday.

■ Strings

Let us revisit the difference between two quaternions squared, as worked out in the section of analysis. A quaternion has 4 degrees of freedom, so it can be represented by 4 real numbers:

$$q = (a_0, a_1, a_2, a_3)$$

Taking the difference between two quaternions is only a valid operation if they share the same basis. Work with defining the derivative with respect to a quaternion has required that a change in the scalar be equal in magnitude to the sum of changes in the 3−vector (instead of the usual parity with components). These concerns lead to the definition of the difference between two quaternions:

$$dq = \left(da_0\, e_0,\ da_1 \frac{e_1}{3},\ da_2 \frac{e_2}{3},\ da_3 \frac{e_3}{3}\right)$$

What type of information must e0, e1, e2, and e3 share in order to make subtraction a valid operation? There is only one basis, so the two events that make up the difference must necessarily be expressed in the same basis. If not, then the standard coordinate transformation needs to be done first. A more subtle issue is that the difference must have the same amount of intrinsic curvature for all three spatial basis vectors. If this is not the case, then it would not longer be possible to do a coordinate transformation using the typical methods. There would be a hidden bump in an otherwise smooth transformation! At this point, I do not yet understand the technical link between basis vectors and intrinsic curvature. I will propose the following relationship between basis vectors because its form suggests a link to intrinsic curvature:

$$-\frac{1}{e_1{}^2} = -\frac{1}{e_2{}^2} = -\frac{1}{e_3{}^2} = e_0{}^2$$

If e0 = 1, this is consistent with Hamilton’s system for 1, i, j, and k. The dimensions for the spatial part are 1/distance^2, the same as intrinsic curvature. This is a flat space, so −1/e1^2 is something like 1 + k. In effect, I am trying to merge the basis vectors of quaternions with tools from topology. In math, I am free to define things as I choose, and if lucky, it will prove useful later on :−)

Form the square of the difference between two quaternion events as defined above:

$$dq^2 = \left(da_0{}^2\, e_0{}^2 + da_1{}^2 \frac{e_1{}^2}{9} + da_2{}^2 \frac{e_2{}^2}{9} + da_3{}^2 \frac{e_3{}^2}{9},\right.$$
$$\left. 2\, da_0\, da_1\, e_0 \frac{e_1}{3},\ 2\, da_0\, da_2\, e_0 \frac{e_2}{3},\ 2\, da_0\, da_3\, e_0 \frac{e_3}{3}\right) =$$

$$= (\text{interval}^2, 3 - \text{string})$$

The scalar is the Lorentz invariant interval of special relativity if e0 = 1.

Why use a work with a powerful meaning in the current physics lexicon for the vector dt dX? A string transforms differently than a spatial 3−vector, the former flipping signs with time, the latter inert. A string will also transform differently under a Lorentz transformation.

The units for a string are time*distance. For a string between two events that have the same spatial location, dX = 0, so the string dt dX is zero. For a string between two events that are simultaneous, dt = 0 so the string is again of zero length. Only if two events happen at different times in different locations will the string be non−zero. Since a string is not invariant under a Lorentz transformation, the value of a string is

We all appreciate the critical role played by the 3−velocity, which is the ratio of dX by dt. Hopefully we can imagine another role as important for the product of these same two numbers.

■ Dimensionless Strings

Imagine some system that happens to create a periodic pattern of intervals and strings (a series of events that when you took the difference between neighboring events and squared them, the results had a periodic pattern). It could happen :−) One might be able to use a collection of sines and cosines to regenerate the pattern, since sines and cosines can do that sort of work. However, the differences would have to first be made dimensionless, since the infinite series expansion for such transcendental functions would not make sense. The first step is to get all the units to be the same, using c. Let a0 have units of time, and a1, a2, a3 have units of space. Make all components have units of time:

$$dq^2 = \left(da_0{}^2\, e_0{}^2 + da_1{}^2\, \frac{e_1{}^2}{9\, c^2} + da_2{}^2\, \frac{e_2{}^2}{9\, c^2} + da_3{}^2\, \frac{e_3{}^2}{9\, c^2},\right.$$
$$\left. 2\, da_0\, da_1\, e_0\, \frac{e_1}{3\, c},\ 2\, da_0\, da_2\, e_0\, \frac{e_2}{3\, c},\ 2\, da_0\, da_3\, e_0\, \frac{e_3}{3\, c}\right)$$

Now the units are time squared. Use a combination of 3 constants to do the work of making this dimensionless.

$$\frac{1}{G} \to \frac{\text{mass time}^2}{\text{distance}^3} \quad \frac{1}{h} \to \frac{\text{time}}{\text{mass distance}^2} \quad c^5 \to \frac{\text{distance}^5}{\text{time}^5}$$

The units for the product of these three numbers are the reciprocal of time squared. This is the same as the reciprocal of the Planck time squared, and in units of seconds is 5.5x10^85s^−2. The symbols needed to make the difference between two events dimensionless are simple:

$$dq^2 = \frac{c^5}{G\, h} \left(da_0{}^2\, e_0{}^2 + da_1{}^2\, \frac{e_1{}^2}{9\, c^2} + da_2{}^2\, \frac{e_2{}^2}{9\, c^2} + da_3{}^2\, \frac{e_3{}^2}{9\, c^2},\right.$$
$$\left. 2\, da_0\, da_1\, e_0\, \frac{e_1}{3\, c},\ 2\, da_0\, da_2\, e_0\, \frac{e_2}{3\, c},\ 2\, da_0\, da_3\, e_0\, \frac{e_3}{3\, c}\right)$$

As far as the units are concerned, this is relativistic (c) quantum (h) gravity (G). Take this constants to zero or infinity, and the difference of a quaternion blows up or disappears.

■ Behaving Like a Relativistic Quantum Gravity Theory

Although the units suggest a possible relativistic quantum gravity, it is more important to see that it behaves like one. Since this unicorn of physics has never been seen I will present 4 cases which will show that this equation behaves like that mysterious beast!

Consider a general transformation T that brings the difference between two events dq into dq'. There are four cases for what can happen to the interval and the string between these two events under this general transformation.

Case 1: Constant Intervals and Strings

$$T: dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) = \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) = \mathrm{vector}(dq'^2)$$

This looks simple, but there is no handle on the overall sign of the 4−dimensional quaternion, a smoke signal of O(4). Quantum mechanics is constructed around dealing with phase ambiguity in a rigorous way. This issue of ambiguous phases is true for all four of these cases.

Case 2: Constant Intervals

$$T: dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) = \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) \neq \mathrm{vector}(dq'^2)$$

Case 2 involves conserving the Lorentz invariant interval, or special relativity. Strings change under such a transformation, and this can be used as a measure of the amount of change between inertial reference frames.

Case 3: Constant Strings

$$T: dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) \neq \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) = \mathrm{vector}(dq'^2)$$

Case 3 involves conserving the quaternion string, or general relativity. Intervals change under such a transformation, and this can be used as a measure of the amount of change between non−inertial reference frames. All that is required to make this simple but radical proposal consistent with experimental tests of general relativity is the following:

$$1 - 2\frac{GM}{c^2 R} = -\frac{1}{e_1{}^2} = -\frac{1}{e_2{}^2} = -\frac{1}{e_3{}^2} = e_0{}^2$$

The string, because it is the product of e0 e1, e0 e2, and e0 e3, will not be changed by this. The phase of the string may change here, since this involves the root of the squared basis vectors. The interval depends directly on the squares of the basis vectors (I think of this as being 1+/− the intrinsic curvature, but do not know if that is an accurate technical assessment). This particular value regenerates the Schwarzschild solution of general relativity.

Case 4: No Constants

$$T : dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) \neq \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) \neq \mathrm{vector}(dq'^2)$$

In this proposal, changes in the reference frame of an inertial observer are logically independent from changing the mass density. The two effects can be measured separately. The change in the length−time of the string will involve the inertial reference frame, and the change in the interval will involve changes in the mass density.

■ The Missing Link

At this time I do not know how to use the proposed unified field equations discussed earlier to generate the basis vectors shown. This will involve determining the precise relationship between intrinsic curvature and the quaternion basis vectors.

Answering Prima Facie Questions in Quantum Gravity Using Quaternions

(Note: this was a post sent to the newsgroup sci.physics.research June 28, 1998)

Chris Isham's paper "Prima Facie Questions in Quantum Gravity" (gr−qc/9310031, October, 1993) details the structure required of any approach to quantum gravity. I will use that paper as a template for this section, noting the highlights (but please refer to this well−written paper for details). Wherever appropriate, I will point out how using quaternions in quantum gravity fits within this superstructure. I will argue that all the technical parts required are all ready part of quaternion mathematics. These tools are required to calculate the smallest norm between two worldlines, which may form a new road to quantum gravity.

■ What Is Quantum Gravity?

Isham sorts the approaches to quantum gravity into four groups. First, there is the classical approach. This begins with Einstein's general relativity. Systematically substitute self−adjoint operators for classical terms like energy and momentum. This gets further subdivided into the 'canonical' scheme where spacetime is split into time and space−−Ashtekar's work−−and a covariant formulation, which is believed to be perturbatively non−renormalizable.

The second approach takes quantum mechanics and transforms it into general relativity. Much less effort has gone in this direction, but there has been work done by Haag.

The third angle has general relativity as the low energy limit of ideas based in conventional quantum mechanics. Quantum gravity dominates the world on the scale of Plank time, length, or energy, a place where only calculations can go. This is where superstring theory lives.

The fourth possibility involves a radical new perspective, where general relativity and quantum mechanics are only different applications of the same mathematical structure. This would require a major "retooling". People with the patience to have read many of my posts (even if not followed :−) know this is the task facing work with quaternions. Replace the tools for doing special relativity−−4−vectors, metrics, tensors, and groups−−with quaternions that preserve the scalar of a squared quaternion. Replace the tools for deriving the Maxwell equations−−4−potentials, metrics, tensors, and groups−−by quaternion operators acting on quaternion potentials using combinations of commutators and anticommutators. It remains to be shown whether quaternions also have the structure required for a quantum gravity theory.

■ Why Do We Study Quantum Gravity?

Isham gives six reasons: the inability to calculate using perturbation theory a correction for general relativity, singularities, quantum cosmology (particularly the Big Bang), Hawking radiation, unification of particles, and the possibility of radical change. This last reason could be a lot of fun, and it is the reason to read this post :−)

■ What Are Prima Facie Questions?

The first question raised by Isham is the relation between classical and quantum physics. Physics with quaternions has a general guide. Consider two arbitrary quaternions, q and q'. The classical distance between them is the interval.

$$\left((t,\ \vec{X}) - \left(t',\ \vec{X'}\right)\right)^2 = \left(dt^2 - d\vec{X}.d\vec{X},\ 2\,dt\,d\vec{X}\right)$$

This involves retooling, because the distance also includes a 3−vector. There is nothing inherently wrong with this vector, and it certainly could be computed with standard tools. To be complete, measure the difference between two quaternions with a quaternion containing the usual invariant scalar interval and a covariant 3−vector. To distinguishing collections of events that are lightlike separated where the interval is zero, use the 3−vector which can be unique. Never discard useful information!

Quantum mechanics involves a Hilbert space. Quaternions can be used to form an inner−product space. The norm of the difference between q and q' is

$$\left((t,\ \vec{X}) - \left(t',\ \vec{X'}\right)\right)^* \left((t,\ \vec{X}) - \left(t',\ \vec{X'}\right)\right) =$$
$$\left(dt^2 + d\vec{X}.d\vec{X},\ \vec{0}\right)$$

The norm can be used to build all the equipment expected of a Hilbert space, including the Schwarz and triangle inequalities. The uncertainty principle can be derived in the same way as is done with the complex−valued wave function.

I call q q' a Grassman product (it has the cross product in it) and q* q' the Euclidean product (it is a Euclidean norm if q = q'). In general, classical physics involves Grassman products and quantum mechanics involves Euclidean products of quaternions.

Isham moves from big questions to ones focused on quantum gravity. Which classical spacetime concepts are needed? Which standard parts of quantum mechanics are needed? Should particles be united? With quaternions, all these concepts are required, but the tools used to build them morph and become unified under one algebraic umbrella.

Isham points out the difficulty of clearly marking a boundary between theories and fact. He writes:

"...what we call a 'fact' does not exist without some theoretical schema for organizing experimental and experiential data; and, conversely, in constructing a theory we inevitably impose some prior idea of what we mean by a fact."

My structure is this: the description of events in spacetime using the topological algebraic field of quaternions is physics.

■ Current Research Programs in Quantum Gravity

There is a list of current approaches to quantum gravity. This is solid a description of the family of approaches being used, circa 1993. See the text for details.

■ Prima Facie Questions in Quantum Gravity

Isham is concerned with the form of these approaches. He writes:

"I mean (by background structure) the entire conceptual and structural framework within whose language any particular approach is couched. Different approaches to quantum gravity differ significantly in the frameworks they adopt, which causes no harm−−indeed the selection of such a framework is an essential pre−requisite for theoretical research−−provided the choice is made consciously."

My framework was stated explicitly above, but it literally does not appear on the radar screen of this discussion of quantum gravity. Moments later comes this comment:

"In using real or complex numbers in quantum theory we are arguably making a prior assumption about the continuum nature of space."

This statement makes a hidden assumption, that quaternions do not belong on a list that includes real and complex numbers. Quaternions have the same continuum properties as the real and complex numbers. The important distinction is that quaternions do not commute. This property is shared by quantum mechanics so it should not banish quaternions from the list. The omission reflects the history of work in the field, not the logic of the mathematical statement.

General relativity may force non−linearity into quantum theory, which require a change in the formalism. It is easy to write non−linear quaternion functions. Near the end of this section I will do that in an attempt to find the shortest norm in spacetime which happens to be non−linear.

Now we come to the part of the paper that got me really excited! Isham described all the machinery needed for classical general relativity. The properties of quaternions dovetail the needs perfectly. I will quote at length, since this is helpful for anyone trying to get a handle on the nature of general relativity.

"The mathematical model of spacetime used in classical general relativity is a differentiable manifold equipped with a Lorentzian metric. Some of the most important pieces of substructure underlying this picture are illustrated in Figure 1.

The bottom level is a set M whose elements are to be identified with spacetime ’points’ or ’events’. This set is formless with its only general mathematical property being the cardinal number. In particular, there are no relations between the elements of M and no special way of labeling any such element.

The next step is to impose a topology on M so that each point acquires a family of neighborhoods. It now becomes possible to talk about relationships between point, albeit in a rather non−physical way. This defect is overcome by adding the key of all standard views of spacetime: the topology of M must be compatible with that of a differentiable manifold. A point can then be labeled uniquely in M (at least locally) by giving the values of four real numbers. Such a coordinate system also provides a more specific way of describing relationships between points of M, albeit not intrinsically in so far as these depend on which coordinate systems are chosen to cover M.

In the final step a Lorentzian metric g is placed on M, thereby introducing the ideas of the length of a path joining two spacetime points, parallel transport with respect to a Riemannian connection, causal relations between pairs of points etc. There are also a variety of possible intermediate steps between the manifold and Lorentzian pictures; for example, as signified in Figure 1, the idea of causal structure is more primitive than that of a Lorentzian metric."

My hypothesis to treat events as quaternions lends more structure than is found in the set M. Specifically, Pontryagin proved that quaternions are a topological algebraic field. Each point has a neighborhood, and limit processes required for a differentiable manifold make sense. Label every quaternion event with four real numbers, using whichever coordinate system one chooses. Earlier in this section I showed how to calculate the Lorentz interval, so the notion of length of a path joining two events is always there. As described by Isham, spacetime structure is built up with care from four unrelated real numbers. With quaternions as events, spacetime structure is the observed properties of the mathematics, inherited by all quaternion functions.

Much work in quantum gravity has gone into viewing how flexible the spacetime structure might be. The most common example involves how quantum fluctuations might effect the Lorentzian metric. Physicists have tried to investigate how such fluctuation would effect every level of spacetime structure, from causality, to the manifold to the topology, even the set M somehow.

None of these avenues are open for quaternion work. Every quaternion equation inherits this wealth of spacetime structure. It is the family quaternion functions are born in. There is nothing to stop combining Grassman and Euclidean products, which at an abstract level, is the way to merge classical and quantum descriptions of collections of events. If a non−linear quaternion function can be defined that is related to the shortest path through spacetime, the cast required for quantum gravity would be complete.

According to Isham, causal structure is particularly important. With quaternions, that issue is particularly straightforward. Could event q have caused q’? Take the difference and square it. If the scalar is positive, then the relationship is timelike, so it is possible. Is it probable? That might depend on the 3−vector, which could be more likely if the vector is small (I don’t understand the details of this suggestion yet). If the scalar is zero, the two have a lightlike relationship. If the scalar is negative, then it is spacelike, and one could not have caused the other.

This causal structure also applies to quaternion potential functions. For concreteness, let q(t) = cos(pi t (2i + 3j + 4k)) and q’(t) = sin(pi t (5i − .1j + 2k). Calculate the square of the difference between q and q’. Depending on the particular value of t, this will be positive, negative or zero. The distance vectors could be anywhere on the map. Even though I don’t know what these particular potential functions represent, the causal relationship is easy to calculate, but is complex and not trivial.

■ The Role of the Spacetime Diffeomorphism Group Diff(M)

Isham lets me off the hook, saying "...[for type 3 and 4 theories] there is no strong reason to suppose that Diff(M) will play any fundamental role in [such] quantum theory." He is right and wrong. My simple tool collection does not include this group. Yet the concept that requires this idea is essential. This group is part of the machinery that makes possible causal measurements of lengths in various topologies. Metrics change due to local conditions. The concept of a flexible, causal metric must be preserved.

With quaternions, causality is always found in the scalar of the square of the difference. For two events in flat spacetime, that is the interval. In curved spacetime, the scalar of the square is different, but it still is either positive, negative or zero.

■ The Problem of Time

Time plays a different role in quantum theory and in general relativity. In quantum, time is treated as a background parameter since it is not represented by an operator. Measurements are made at a particular time. In classical general relativity in curved spacetime, there are many possible metrics which might work, but no way to pick the appropriate one. Without a clear definition of measurement, the definition is non−physical. Fixing the metric cannot be done if the metric is subject to quantum fluctuations.

Isham raises three questions:

"How is the notion of time to be incorporated in a quantum theory of gravity?

Does it play a fundamental role in the construction of the theory or is it a ’phenomenological’ concept that applies, for example, only in some coarse−grained, semi−classical sense?

In the latter case, how reliable is the use at a basic level of techniques drawn from standard quantum theory?"

Three solutions are noted: fix the background causal structure, locate events within functionals of fields, or make no reference to time.

With quaternions, time plays a central role, and is in fact the center of the matrix representation. Time is isomorphic to the real numbers, so it forms a totally ordered sub−field of the quaternions. It is not time per se, but the location of time within the event quaternion (t, x i, y j, z k) that gives time its significance. The scalar slot can be held by energy (E, px i, py j, pz k), the tangent of spacetime, by the interval of classical physics (t^2 − x^2 − y^2 − z^2, 2 tx i, 2 ty j, 2 tz k) or the norm of quantum mechanics (t^2 + x^2 + y^2 + z^2, 0, 0, 0). Time, energy, intervals, norms,...they all can take the same throne isomorphic to the real numbers, taking on the properties of a totally ordered set within a larger, unordered framework. Events are not totally ordered, but time is. Energy/momenta are not totally ordered, but energy is. Squares of events are not totally ordered, but intervals are. Norms are totally ordered and bounded below by zero.

Time is the only element in the scalar of an event. Time appears in different guises for the scalars of energy, intervals and norms. The richness of time is in the way it weaves through these other scalars, sharing the center in different ways with space.

■ Approaches to Quantum Gravity

Isham surveys the field. At this point I think I’ll just explain my approach. It is based on a concept from general relativity. A painter falling from a ladder travels along the shortest path through spacetime. How does one go about finding the shortest path? In Euclidean 3−space, that involves the triangle inequality. A proof can be done using quaternions if the scalar is set to zero. That proof can be repeated with the scalar set free. The result is the shortest distance through spacetime, or gravity, according to general relativity.

What is the shortest distance between two points A and B in Euclidean 3−space?

$$A = (0,\ ax,\ ay,\ az)$$

$$B = (0,\ bx,\ by,\ bz)$$

What is the shortest distance between two worldlines A(t) and B(t) in spacetime?

$$A(t) = (t,\ ax(t),\ ay(t),\ az(t))$$

$$B(t) = (t,\ bx(t),\ by(t),\ bz(t))$$

The Euclidean 3−space question is a special case of the worldline question. The same proof of the triangle inequality answers both questions. Parameterize the norm N(k) of the sum of A(t) and B(t).

$$N(k) = (A + k B)^* (A + k B)$$

$$= A^* A + k (A^* B + B^* A) + k^2 B^* B$$

Find the extremum of the parameterized norm.

$$\frac{dN}{dk} = 0 = A^* B + B^* A + 2 k B^* B$$

The extremum is a minimum

$$\frac{d^2 N}{dk^2} = 2 B^* B \geq 0$$

The minimum of a quaternion norm is zero. Plug the extremum back into the first equation.

$$0 \le A^* A - \frac{(A^* B + B^* A)^2}{2 B^* B} + \frac{(A^* B + B^* A)^2}{4 B^* B}$$

Rearrange.

$$(A^* B + B^* A)^2 \le 4 A^* A B^* B$$

Take the square root.

$$A^* B + B^* A \le 2 \sqrt{A^* A B^* B}$$

Add the norm of A and B to both sides.

$$A * A + A^* B + B^* A + B^* B \le A^* A + 2 \sqrt{A^* A B^* B} + B^* B$$

Factor.

$$N (A + B) = (A + B)^* (A + B) \le \left(\sqrt{A^* A} + \sqrt{B^* B}\right)^2$$

The norm of the worldline of A plus B is less than the norm of A plus the norm of B.

List the mathematical structures required. To move the triangle inequality from Euclidean 3−space to worldlines required the inclusion of the scalar time component of quaternions. The proof required differentiation to find the minimum. The norm is a Euclidean product, which plays a central role in quaternion quantum mechanics. Doubling A or B does not double the norm of the sum due to cross terms, so the minimal function is not linear.

To address a question raised by general relativity with quaternions required all the structure Isham suggested except causality using the Grassman product. The above proof could be repeated using Grassman products. The only difference would be that the extremum would be an interval which can be positive, negative or zero (a minimum, a maximum or an inflection point).

■ Certainty Is Seven for Seven

I thought I'd end this long section with a personal story. At the end of my college days, I started drinking heavily. Not alcohol, soda. I'd buy a Mellow Yellow and suck it down in under ten seconds. See, I was thirsty. Guzzle that much soda, and, well, I also had to go to the bathroom, even in the middle of the night. I was trapped in a strange cycle. Then I noticed my tongue was kind of foamy. Bizarre. I asked a friend with diabetes what the symptoms of that disease were. She rattled off six: excessive thirst, excessive urination, foamy tongue, bad breath, weight loss, and low energy. I concluded on the spot I had diabetes. She said that I couldn't be certain. Six for six is too stringent a match, and I felt very confident I had this chronic illness. I got the seventh later when she tested my blood glucose on her meter and it was off−scale. She gave me sympathy, but I didn't feel at all sorry for myself. I wanted facts: how does this disease work and how do I cope?

Nothing was made official until I visited the doctor and he ran some tests. The doctor's prescription got me access to the insulin I could no longer produce. It was, and still is today, a lot of work to manage the disease.

When I look at Isham's paper, I see six constraints on the structure of any approach to quantum gravity: events are sets of 4 numbers, events have topological neighborhoods, they live on differential manifolds, there is one of the three types of causal relationships between all events, the distance between events is the interval whose form can vary and a Hilbert space is required for quantum mechanics. Quaternions are six for six. The seventh match is the non−linear shortest norm of spacetime. I have no doubt in the diagnosis that the questions in quantum gravity will be answered with quaternions. Nothing here is official. There are many test that must be passed. I don’t know when the doctor will show up and make it official. It will take a lot of work to manage this solution.

Length in Curved Spacetime

■ The Affine Parameter of General Relativity

The affine parameter is defined in Misner, Thorne and Wheeler as a multiple of the proper time plus a displacement.

$$\lambda = a\,\tau + b$$

The affine parameter is used to determine length in curved spacetime. In this section, the length of a quaternion in curved spacetime will be analyzed. Under certain approximations, this length will depend on the square of the affine parameter, but the two measures are slightly different.

■ Length in Flat Spacetime

Calculating the square of the interval between two events in flat spacetime was straightforward: take the difference between two quaternions and square it.

$$L_{flat} = (q - q')^2 = \left(dt^2 - d\vec{X}^2,\ 2\,dt\,d\vec{X}\right)$$

The first term is the square of the interval. Spacetime is flat in the sense that the first term is exactly like the Minkowski metric in spacetime. There are quaternions which preserve the interval, and those quaternions were used to solve problems in special relativity.

Although not important in this context, it is significant that the value of the vector portion depends upon the observer. This gives a way to distinguish between various frequencies of light for example.

■ Length in Curved Spacetime

Consider if the origin is located at two different locations in spacetime. Characterize each origin as a quaternion, calling the o and o'. In flat spacetime, the two origins would be identical. Calculate the interval as done above, but account for the change in the origin.

$$L_{curved} = ((q + o) - (q' + o'))^2$$

$$= \left(d\,(t + t_o)^2 - d\,(\vec{X} + \vec{X}_o)^2,\ 2\,d\,(t + t_o)\,d\,(\vec{X} + \vec{X}_o)\right)$$

Examine the first term more closely by expanding it.

$$\left(dt^2 - d\vec{X}^2\right) + \left(dt_o^2 - d\vec{X}_o^2\right) + 2\,dt\,dt_o - 2\,d\vec{X}\,d\vec{X}_o$$

The length in curved spacetime is the square of the interval (invariant under a boost) between the two origins, plus the square of the interval between the two events, plus a cross term, which will not be invariant under a boost. The length is symmetric under exchange of the event with the origin translation.

L curved looks similar to the square of the affine parameter:

$$\lambda^2 = b^2 + 2\,a\,b\,\tau + a^2\,\tau^2$$

In this case, b^2 is the origin interval squared and a = 1. There is a difference in the cross terms. However, in the small curvature limit, delta to >> delta Xo, so tau ~ delta to. Under this approximation, the square of the affine parameter and L curved are the same.

For a strong gravitational field, L curved will be different than the square of the affine parameter. The difference will be solely in the nature of the cross term. In general relativity, b and tau are invariant under a boost. For L curved, the cross term should be covariant. Whether this has any effects that can be measured needs to be explored.

There exist quaternions which preserve L curved because quaternions are a field (I haven't found them yet because the math is getting tough at this point!) It is my hope that those quaternions will help solve problems in general relativity, as was the case in special relativity.

■ Implications

A connection to the curved geometry of general relativity was sketched. It should be possible to solve problems with this "curved" measure. As always, all the objects employed were quaternions. Therefore any of the previously outline techniques should be applicable. In particular, it will be fun in the future to think about things like

$$((q + o) - (q' + o'))^* ((q + o) - (q' + o'))$$

$$= \left(d(t + t_o)^2 + d(\vec{X} + \vec{X}_o)^2,\ 2\,d(t + t_o)\,d(\vec{X} + \vec{X}_o)\right)$$

$$= \left(\left(dt^2 - d\vec{X}^2\right) + \left(dt_o^2 - d\vec{X}_o^2\right) + 2\,dt\,dt_o + 2\,d\vec{X}\,d\vec{X}_o,\ \ ...\right)$$

which could open the door to a quantum approach to curvature.

A New Idea for Metrics

In special relativity, the Minkowski metric is used to calculate the interval between two spacetime intervals for inertial observers. Einstein recognized that inertial observes were "special", a unique class. Therefore he set out to understand what was the most general notion for transformations and metrics. This lead to his study of Riemannian geometry, and eventually to general relativity. In this section I shall start from the Lorentz invariant interval using quaternions, then try to generalize this approach using a different way which might prove compatible with quantum mechanics.

For the physics of gravity, general relativity (GR) makes the right predictions of all experimental tests conducted to date. For the physics of atoms, quantum mechanics (QM) makes the right predictions to an even high degree of precision. The problem of building a quantum theory of gravity (QG) hides between general relativity and quantum mechanics. General relativity deals with the measurements of intervals in curved spacetime, special relativity (SR) being adapted to work in flat space. Quantum mechanics is used to calculate the norms of wave functions in a flat linear space. A quantum gravity theory will be used to calculate norms of wave functions in curved space.

		Measurement	
		interval	norm
diff.	flat	SR	QM
geo.	curved	GR	QG

This chart suggests that the form of measurement (interval/norm) should be independent of differential geometry (flat/curved). That will be the explicit goal of this section.

Quaternions come with a metric, a means of taking 4 numbers and returning a scalar. Hamilton defined the roles like so:

$$\vec{i}^2 = \vec{j}^2 = \vec{k}^2 = -1 \quad \vec{i}\,\vec{j}\,\vec{k} = -1$$

The scalar result of squaring a differential quaternion in the interval of special relativity:

$$\text{scalar}\left((dt,\ d\vec{X})^2\right) = dt^2 - d\vec{X}.d\vec{X}$$

How can this be generalized? It might seem natural to explore variations on Hamilton’s rules shown above. Riemannian geometry uses that strategy. When working with a field like quaternions, that approach bothers me because Hamilton’s rules are fundamental to the very definition of a quaternion. Change these rules and it may not be valid to compare physics done with different metrics. It may cause a compatibility problem.

Here is a different approach which generalizes the scalar of the square while being consistent with Hamilton’s rules.

$$\text{interval}^2 = \text{scalar}\,(g\,dq\,g\,dq)$$

$$\text{if } g = \left(1, \vec{0}\right),$$

$$\text{then interval}^2 = dt^2 - d\vec{X}.d\vec{X}$$

If g is the identity matrix. Then then result is the flat Minkowski interval. The quaternion g could be anything. What if g = i? (what would you guess, I was surprised :−)

$$\text{scalar}\left(\left((0, 1, 0, 0)(t, x, y, z)\right)^2\right) =$$

$$= \left(-t^2 + x^2 - y^2 - z^2, \vec{0}\right)$$

Now the special direction x plays the same role as time! Does this make sense physically? Here is one interpretation. When g=1, a time−like interval is being measured with a wristwatch. When g=i, a space−like interval along the x axis is being measured with a meter stick along the x axis.

Examine the most general case, where small letters are scalar, and capital letters are 3−vectors:

$$\text{interval}^2 = \text{scalar}\left((g, \vec{G})(dt, d\vec{X})(g, \vec{G})(dt, d\vec{X})\right) =$$

$$= g^2 (dt^2 - d\vec{X}.d\vec{X}) - 4\, g\, dt\, \vec{G}.d\vec{X} +$$
$$(\vec{G}.d\vec{X})^2 - dt^2\, d\vec{G}.d\vec{G} - (\vec{G} \times d\vec{X}).(\vec{G} \times d\vec{X}) =$$

In component form...

$$= (+ g^2 - Gx^2 - Gy^2 - Gz^2)\, dt^2 +$$
$$+ (- g^2 + Gx^2 - Gy^2 - Gz^2)\, dx^2 +$$
$$+ (- g^2 - Gx^2 + Gy^2 - Gz^2)\, dy^2$$
$$+ (- g^2 - Gx^2 - Gy^2 + Gz^2)\, dz^2 +$$
$$- 4\, g\, Gx\, dt\, dx - 4\, g\, Gy\, dt\, dy - 4\, g\, Gz\, dt\, dz$$
$$+ 4\, Gx\, Gy\, ds\, dy + 4\, Gx\, Gz\, dx\, dz + 4\, Gy\, Gz\, dy\, dz$$

This has the same combination of ten differential terms found in the Riemannian approach. The difference is that Hamilton’s rule impose an additional structure.

I have not yet figured out how to represent the stress tensor, so there are no field equations to be solved. We can figure out some of the properties of a static, spherically−symmetric metric. Since it is static, there will be no terms with the deferential element dt dx, dt dy, or dt dz. Since it is spherically symmetric, there will be no terms of the form dx dy, dx dz, or dy dz. These constraints can both be achieved if Gx = Gy = Gz = 0. This leaves four differential equations.

Here I will have to stop. In time, I should be able to figure out quaternion field equations that do the same work as Einstein’s field equations. I bet it will contain the Schwarzschild solution too :−) Then it will be easy to create a Hilbert space with a non−Euclidean norm, a norm that is determined by the distribution of mass−energy. What sort of calculation to do is a mystery to me, but someone will get to that bridge...

The Gravitational Redshift

Gravitational redshift experiments are tests of conservation of energy in a gravitational potential. A photon lower in a gravitational potential expends energy to climb out, and this energy cost is seen as a redshift. In this section, the difference between weak gravitational potentials will be calculated and shown to be consistent with experiment. Quaternions are not of much use here because energy is a scalar, the first term of a quaternion that is a scalar multiple of the identity matrix.

■ The Pound and Rebka Experiment

The Pound and Rebka experiment used the Mossbauer effect to measure a redshift between the base and the top of a tower at Harvard University. The relevant potentials are

$$\phi_{\text{tower}} = \frac{G\,M}{r + h};$$

$$\phi_{\text{base}} = \frac{G\,M}{r};$$

The equivalence principle is used to transform the gravitational potential to a speed (this only involves dividing phi by the constant c^2).

$$\beta_{\text{tower}} = \frac{G\,M}{c^2\,(r + h)};$$

$$\beta_{\text{base}} = \frac{G\,M}{c^2\,r};$$

Now the problem can be viewed as a relativistic Doppler effect problem. A redshift in a frequency is given by

$$\nu' == (\gamma[\beta] + \beta\,\gamma[\beta])\,\nu_o$$

For small velocities, the Doppler effect is

$$\text{Series}[\gamma[\beta] + \beta\,\gamma[\beta], \{\beta, 0, 1\}]$$

$$= 1 + \beta + O[\beta]^2$$

The experiment measured the difference between the two Doppler shifts.

$$\text{Series}[((1 + \beta_{\text{tower}}) - (1 + \beta_{\text{base}}))\,\nu_o, \{h, 0, 1\}]$$

$$= -\frac{G\,M\,\nu_o\,h}{c^2\,r^2} + O[h]^2$$

Or equivalently,

$$\nu' = g h \nu_o$$

This was the measured effect.

■ Escape From a Gravitational Potential

A photon can escape from a star and travel to infinity (or to us, which is a good approximation). The only part of the previous calculation that changes is the limit in the final step.

$$\text{Limit}[((1+\beta_{tower})-(1+\beta_{base}))\,\nu_o,\ h \to \text{Infinity}]$$

$$= -\frac{G M \nu_o}{c^2 r}$$

This shift has been observed in the spectral lines of stars.

■ Clocks at different heights in a gravitational field

C. O. Alley conducted an experiment which involved flying an atomic clock at high altitude and comparing it with an atomic clock on the ground. This is like integrating the redshift over the time of the flight.

$$\int_0^t -\frac{G M h}{c^2 r^2}\,dt = -\frac{G h M t}{c^2 r^2}$$

This was the measured effect.

■ Implications

Conservation of energy involves the conservation of a scalar. Consequently, nothing new will happen by treating it as a quaternion. The approach used here was not the standard one employed. The equivalence principle was used to transform the problem into a relativistic Doppler shift effect. Yet the results are no different. This is just part of the work to connect quaternions to measurable effects of gravity.

■ References

For the Pound and Rebka experiment, and escape:

Misner, Thorne, and Wheeler, Gravitation, 1970.

For the clocks at different heights:

Quantum optics, experimental gravitation and measurement theory, Ed. P. Meystre, 1983 (also mentioned in Taylor and Wheeler, Spacetime Physics, section 4.10)

A Brief Summary of Important Laws in Physics Written as Quaternions

Summary

Each of the following laws of physics are generated by quaternion operators acting on the appropriate quaternion−valued functions. The generators of these common laws often provide insight.

■ Classical Mechanics

Newton’s 2nd Law for an Inertial Reference Frame in Cartesian Coordinates

$$A = \left(\frac{d}{dt}, \vec{0}\right) \left(1, \dot{\vec{R}}\right) = \left(0, \ddot{\vec{R}}\right)$$

Newton’s 2nd Law in Polar Coordinates for a Central Force in a Plane

$$A = (\mathrm{Cos}[\theta],\ 0,\ 0,\ -\mathrm{Sin}[\theta]) \left(\frac{d}{dt}, \vec{0}\right)^2 (t,\ R\,\mathrm{Cos}[\theta],\ R\,\mathrm{Sin}[\theta],\ 0) =$$

$$= \left(0,\ \frac{L^2}{m^2\, R^3} + \ddot{\vec{R}},\ \frac{2\, L\, \dot{\vec{R}}}{m\, R^2},\ 0\right)$$

Newton’s 2nd Law in a Noninertial, Rotating Frame

$$A = \left(\frac{d}{dt},\ \vec{\omega}\right)\ \left(-\ \vec{\omega}.\vec{R}\ ,\ \dot{\vec{R}}\ +\ \vec{\omega}\ \mathrm{x}\ \vec{R}\right)\ =$$

$$= \left(-\ \dot{\vec{\omega}}.\vec{R}\ ,\ \ddot{\vec{R}} + 2\,\vec{\omega}\ \mathrm{x}\ \dot{\vec{R}} + \dot{\vec{\omega}}\ \mathrm{x}\ \vec{R} - \vec{\omega}.\vec{R}\,\vec{\omega}\right)$$

The Simple Harmonic Oscillator (SHO)

$$\left(\frac{d}{dt},\ \vec{0}\right)^2\ (0,\ x,\ 0,\ 0)\ +\ \left(0,\ \frac{k}{m}\,x,\ 0,\ 0\right)\ =$$

$$\left(0,\ \frac{d^2\,x}{dt^2} + \frac{k\,x}{m},\ 0,\ 0\right) = 0$$

The Damped Simple Harmonic Oscillator

$$\left(\frac{d}{dt},\ \vec{0}\right)^2\ (0,\ x,\ 0,\ 0)\ +$$

$$\left(\frac{d}{dt},\ \vec{0}\right)\ (0,\ b\,x,\ 0,\ 0)\ +\ \left(0,\ \frac{k}{m}\,x,\ 0,\ 0\right)\ =$$

$$= \left(0,\ \frac{d^2\,x}{dt^2} + \frac{b\,d\,x}{dt} + \frac{k\,x}{m},\ 0,\ 0\right) = 0$$

The Wave Equation

$$\left(\frac{d}{v\,dt},\ \frac{d}{dx},\ 0,\ 0\right)^2\ (0,\ 0,\ f[t\,v\ +\ x],\ 0)\ =$$

$$= \left(0,\ 0,\ \left(-\frac{d^2}{dx^2} + \frac{d^2}{dt^2\,v^2}\right) f[t\,v + x],\ \frac{2\,d^2\,f[t\,v + x]}{dt\,dx\,v}\right)$$

The third term is the one dimensional wave equation. The forth term is the instantaneous power transmitted by the wave.

A Force Is Conservative If The Curl Is Zero

$$\mathrm{odd}\ \left(\left(\frac{d}{dt},\ \vec{\nabla}\right),\ \vec{F}\right) = 0$$

A Force Is Conservative If There Exists a Potential Function for the Force

$$F = \left(\frac{d}{dt}, \vec{\nabla}\right)(\phi, \vec{0})$$

A Force Is Conservative If the Line Integral of Any Closed Loop Is Zero

$$\oint F\,dt = 0$$

A Force Is Conservative If the Line Integral Along Different Paths Is the Same

$$\left(\underset{\text{path 1}}{\oint}\right) F\,dt = \left(\underset{\text{path 2}}{\oint}\right) F\,dt$$

■ Special Relativity

Rotations and Dilations Create the Lorentz Group

$$q' = q + (\gamma - 1)\frac{\text{even}\left(\text{even}\left(\vec{\nabla}^*, q\right), \vec{\nabla}\right)}{|\vec{\nabla}|^2} + \gamma\,\text{even}\left(\vec{\nabla}^*, q^*\right)$$

An Alternative Algebra for Lorentz Boosts

$$\text{scalar}((t, x, y, z)^2) = \text{scalar}((L(t, x, y, z))^2)$$

For boosts along the x axis...

If t = 0, then

$$L = \gamma(1, \beta, 0, 0)$$

If x = 0, then

$$L = \gamma(1, -\beta, 0, 0)$$

If t = x, then for blueshifts

$$L = \gamma(1 - \beta, 0, 0, 0)$$

For general boosts along the x axis

$$L = (\gamma t^2 + \gamma x^2 - 2\gamma\beta t x + (y^2 + z^2), \gamma\beta(-t^2 + x^2),$$
$$t(\beta\gamma z + y(1-\gamma)) - x(\gamma\beta y + z(1-\gamma)),$$
$$t(\gamma\beta y + z(1-\gamma)) + x(\gamma\beta z + y(1-\gamma)))\,/$$
$$(t^2 + x^2 + y^2 + z^2)$$

■ Electromagnetism

The Maxwell Equations

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{B})\right) + \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{E})\right) =$$
$$\left(-\vec{\nabla}\cdot\vec{B},\ \vec{\nabla} \times \vec{E} + \frac{\partial\vec{B}}{\partial t}\right) = \left(0, \vec{0}\right)$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{B})\right) - \text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (0, \vec{E})\right) =$$
$$\left(\vec{\nabla}\cdot\vec{E},\ \vec{\nabla} \times \vec{B} - \frac{\partial\vec{E}}{\partial t}\right) = 4\pi(\rho, \vec{J})$$

■ Maxwell Written with Potentials

The fields

$$E = \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) = \left(0, -\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\phi\right)$$

$$B = \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right) = (0, \vec{\nabla} \times \vec{A})$$

The field equations

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) +$$
$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right)\right) =$$
$$= \left(-\vec{\nabla}\cdot\vec{\nabla}\times\vec{A},\ \frac{\partial\vec{\nabla}\times\vec{A}}{\partial t} - \vec{\nabla}\times\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}\times\vec{\nabla}\phi\right) =$$
$$\left(-\vec{\nabla}\cdot\vec{B},\ \frac{\partial\vec{B}}{\partial t} + \vec{\nabla}\times\vec{E}\right) = \left(0, \vec{0}\right)$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, \vec{A})\right)\right) -$$

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{vector}\left(\text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right)\right) =$$

$$= \left(-\vec{\nabla}\cdot\vec{\nabla}\phi - \vec{\nabla}\cdot\frac{\partial\vec{A}}{\partial t}, \ \vec{\nabla}\,X\,\vec{\nabla}\,X\,\vec{A} + \frac{\partial^2\vec{A}}{\partial t^2} + \frac{\partial\vec{\nabla}\phi}{\partial t}\right) =$$

$$\left(\vec{\nabla}\cdot\vec{E}, \ \vec{\nabla}\,X\,\vec{B} - \frac{\partial\vec{E}}{\partial t}\right) = 4\pi(\rho, \vec{J})$$

The Lorentz Force

$$\text{odd}\left((\gamma, \ \gamma\vec{\beta}), (0, \vec{B})\right) - \text{even}\left((-\gamma, \ \gamma\vec{\beta}), (0, \vec{E})\right) =$$

$$\left(\gamma\vec{\beta}\cdot\vec{E}, \ \gamma\vec{E} + \gamma\vec{\beta}\,X\,\vec{B}\right)$$

Conservation Laws

The continuity equation

$$\text{scalar}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\vec{\nabla}\cdot\vec{E}, \vec{\nabla}\,X\,\vec{B} - \frac{\partial\vec{E}}{\partial t}\right)\right) =$$

$$\left(\frac{\partial}{\partial t}\vec{\nabla}\cdot\vec{E} - \vec{\nabla}\cdot\frac{\partial\vec{E}}{\partial t} + \vec{\nabla}\cdot\vec{\nabla}\,X\,\vec{B}, \ 0\right) =$$

$$= \text{scalar}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), 4\pi(\rho, \vec{J})\right) = 4\pi\left(\vec{E}\cdot\vec{J} + \frac{\partial\rho}{\partial t}, \ 0\right)$$

Poynting’s theorem for energy conservation.

$$\text{scalar}\left((0, -\vec{E})\left(\vec{\nabla}\cdot\vec{E}, \vec{\nabla}\,X\,\vec{B} - \frac{\partial\vec{E}}{\partial t}\right)\right) =$$

$$\left(\vec{E}\cdot\vec{\nabla}\,X\,\vec{B} - \vec{E}\cdot\frac{\partial\vec{E}}{\partial t}, \ 0\right) =$$

$$\left(-\vec{\nabla}\cdot(\vec{E}\,X\,\vec{B}) - \frac{1}{2}\left(\frac{\partial\vec{E}}{\partial t}\right)^2 - \frac{1}{2}\left(\frac{\partial\vec{B}}{\partial t}\right)^2, \ 0\right)$$

$$= \text{scalar}\left((0, -\vec{E}), 4\pi(\rho, \ \vec{J})\right) = 4\pi(\vec{E}\cdot\vec{J}, \ 0)$$

The Field Tensor F in Different Gauges

The anti−symmetric 2−rank electromagnetic field tensor F

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) - (\phi, \vec{A})\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$\left(0, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

F in the Lorenz gauge.

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) + (\phi, -\vec{A})}{2}\right) -$$

$$\left(\frac{(\phi, \vec{A}) - (\phi, -\vec{A})}{2}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(\frac{\partial \phi}{\partial t} + \vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

F in the Coulomb gauge

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) + \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) - (\phi, -\vec{A})}{4}\right) +$$

$$\left(\frac{(\phi, -\vec{A}) - (\phi, \vec{A})}{4}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(\frac{\partial \phi}{\partial t}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

F in the temporal gauge.

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) - \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{(\phi, \vec{A}) + (\phi, -\vec{A})}{4}\right) -$$

$$\left(\frac{(\phi, -\vec{A}) + (\phi, \vec{A})}{4}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right) =$$

$$= \left(-\vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

F in the light gauge.

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla}.\vec{A}, -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} X \vec{A}\right)$$

The light gauge is one sign different from the Lorenz gauge, but its generator is a simple as it gets.

The Maxwell Equations in the Light Gauge

Note: subsequent work has suggested that the scalar in these equations is part of a unified field theory.

$$\text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right)\right) + \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) =$$

$$= (-\vec{\nabla} \cdot \vec{\nabla} X \vec{A}, -\vec{\nabla} X \vec{\nabla} \phi) = (0, \vec{0})$$

$$\text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{odd}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), (\phi, \vec{A})\right)\right) - \text{even}\left(\left(\frac{\partial}{\partial t}, \vec{\nabla}\right), \text{even}\left(\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right), (\phi, -\vec{A})\right)\right) =$$

$$= \left(\frac{\partial^2 \phi}{\partial t^2} + \vec{\nabla} \cdot \vec{\nabla} \phi, -\frac{\partial^2 \vec{A}}{\partial t^2} + \vec{\nabla} X (\vec{\nabla} X \vec{A}) - \vec{\nabla} \vec{\nabla} \cdot \vec{A}\right) =$$

$$\left(\frac{\partial^2 \phi}{\partial t^2} + \vec{\nabla}^2 \phi, -\frac{\partial^2 \vec{A}}{\partial t^2} - \vec{\nabla}^2 \vec{A}\right) = 4 \pi (\rho, \vec{J})$$

The Stress Tensor of the Electromagnetic Field

$$T^{ik} = \sum_{a=x}^{y,z} \sum_{b=x}^{y,z} \frac{1}{4\pi} \left(\left(\frac{\text{even}(Ua, Ub)}{3} - 1\right) \frac{((0, E)^2 + (0, B)^2)}{2} - \right.$$

$$- \text{even}(E, Ua)\, \text{even}(E, Ub) - \text{even}(B, Ua)\, \text{even}(B, Ub) -$$

$$- \text{even}(\text{odd}(E, B), Ua) - \text{even}(\text{odd}(E, B), Ub) =$$

$$= (-Ex\,Ey - Ex\,Ez - Ey\,Ez - Bx\,By - Bx\,Bz - By\,Bz + Ey\,Bz - Ez\,By + Ez\,Bx - Ex\,Bz + Ex\,By - Ey\,Bx,\ 0) / 2\pi$$

■ Quantum Mechanics

Quaternions in Polar Coordinate Form

$$q = ||q|| \exp[\theta \vec{I}] = q^* q (\text{Cos}[\theta] + \vec{I}\, \text{Sin}[\theta])$$

Multiplying Quaternion Exponentials

$$q\,q' = \{q,\ q'\}^* + \mathrm{Abs}[q,\ q']^*\ \mathrm{Exp}\left[\frac{\pi}{2}\ \frac{[q,\ q']^*}{\mathrm{Abs}[q,\ q']^*}\right]$$

Commutators of Observable Operators

$$[\hat{A},\ \hat{B}]\,q = (\hat{A}\,\hat{B} - \hat{B}\,\hat{A})\,q = -a\,I\,\frac{d\,q}{d\,a} + I\,\frac{d\,a\,q}{d\,a}$$

$$= -a\,I\,\frac{d\,q}{d\,a} + a\,I\,\frac{d\,q}{d\,a} + I\,q\,\frac{d\,a}{d\,a} = I\,q$$

The Uncertainty Principle

$$\frac{[A,\ B]}{2} = \frac{I}{2} \le \delta A^2\ \delta B^2$$

Unifying the Representation of Spin and Angular Momentum

For small rotations:

$$[R_{e_1=0},\ R_{e_2=0}] = 2\,(R_{e_3=0}\,(\Theta^2) - R\,(0))$$

Automorphic Commutator Identities

$$[q,\ q'] = [q^*,\ q'^*] = [q^{*1},\ q'^{*1}]^{*1} = [q^{*2},\ q'^{*2}]^{*2}$$

$$\{q,\ q'\} = \{q^*,\ q'^*\}^* = -\{q^{*1},\ q'^{*1}\}^{*1} = -\{q^{*2},\ q'^{*2}\}^{*2}$$

The Schrödinger Equation

$$\Psi = \mathrm{Exp}\left(\frac{\vec{\nabla}}{\sqrt{\vec{\nabla}.\vec{\nabla}}}\ (\omega\,t - \vec{K}.\vec{X})\right)$$

$$H\,\psi = -i\,\hbar\,\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2\,m}\,\nabla^2\,\psi + V\,(0,\ X)\,\psi$$

The Klein−Gordon Equation

$$\sum_{n=0}^{\infty}\left(\left(\frac{\partial}{\partial t},\ \vec{\nabla}\right)^2 + \left(\frac{\partial}{\partial t},\ -\vec{\nabla}\right)^2 + (E_n,\ \vec{P}_n)^2 + (E_n,\ -\vec{P}_n)^2\right)$$

$$(E_n,\ \vec{P}_n)\,/\,2 =$$

$$= \sum_{n=0}^{\infty} \Big(-\vec{\nabla} \cdot (\vec{\nabla} \, X \, \vec{P}_n) - \vec{\nabla} \cdot \vec{\nabla} \, E_n -$$

$$\vec{P}_n \cdot (\vec{P}_n \, X \, \vec{P}_n) - (\vec{P}_n \cdot \vec{P}_n) \, E_n + E_n{}^3 + \frac{\partial^2 E_n}{\partial t^2},$$

$$\vec{\nabla} \, X \, (\vec{\nabla} \, X \, \vec{P}_n) + \vec{\nabla} \, X \, (\vec{\nabla} \, E_n) + \vec{P}_n \, X \, (\vec{P}_n \, X \, \vec{P}_n) + (\vec{P}_n \, X \, \vec{P}_n) \, E_n - \vec{\nabla}$$

$$(\vec{\nabla} \cdot \vec{P}_n) + \vec{P}_n \, E_n{}^2 - \vec{P}_n \, (\vec{P}_n \cdot \vec{P}_n) + \frac{\partial^2 \vec{P}_n}{\partial t^2} \Big)$$

It takes some skilled staring to assure that this equation contains the Klein−Gordon equation along with vector identities.

Time Reversal Transformations for Intervals

$$(t, \vec{X}) \; -> \; (-t, \vec{X}) = R \, (t, \vec{X})$$

$$R = (-t, \vec{X}) \, (t, \vec{X})^{-1} = (-t^2 + \vec{X}.\vec{X}, \, 2 \, t \, \vec{X}) \, / \, (t^2 + \vec{X}.\vec{X})$$

Classically

$$\text{if } \beta << 1 \text{ then } R \approx \left(-1, \, 2 \, t \, \vec{\beta}\right)$$

$$R = \left(-\frac{\epsilon}{T}, \, 1, \, 0, \, 0\right)$$

■ Gravity

The 3 Fields: g, E & B

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) =$$
$$\left(\dot{\phi} - \vec{\nabla}.\vec{A},\ -\dot{\vec{A}} - \vec{\nabla}\phi + \vec{\nabla} \times \vec{A}\right) = (g, \vec{E} + \vec{B})$$

Field Equations: Almost Maxwell and a Dynamic g

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)(g, \vec{E} + \vec{B}) =$$
$$\left(\dot{g} - \vec{\nabla}.\vec{E} - \vec{\nabla}.\vec{B},\ \dot{\vec{E}} + \vec{\nabla} \times \vec{B} + \dot{\vec{B}} + \vec{\nabla} \times \vec{E} + \vec{\nabla} g\right) =$$
$$4\pi(\rho_g + \rho_e,\ \vec{J}_g + \vec{J}_e)$$

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(g, \vec{E} + \vec{B}) =$$
$$\left(\dot{g} + \vec{\nabla}.\vec{E} + \vec{\nabla}.\vec{B},\ \dot{\vec{E}} - \vec{\nabla} \times \vec{B} + \dot{\vec{B}} - \vec{\nabla} \times \vec{E} - \vec{\nabla} g\right) =$$
$$4\pi(\rho_g + \rho_e,\ \vec{J}_g + \vec{J}_e)$$

Recreating Maxwell

$$\text{Let } U = \left(-\vec{\nabla}.\vec{E} - \vec{\nabla}.\vec{B} + \dot{g},\ \dot{\vec{E}} + \vec{\nabla} \times \vec{B} + \dot{\vec{B}} + \vec{\nabla} \times \vec{E} + \vec{\nabla} g\right)$$

$$W = \left(\vec{\nabla}.\vec{E} + \vec{\nabla}.\vec{B} + \dot{g},\ \dot{\vec{E}} - \vec{\nabla} \times \vec{B} + \dot{\vec{B}} - \vec{\nabla} \times \vec{E} - \vec{\nabla} g\right)$$

$$\text{Mirror}((W + U)/2) + (W - U)^*/2 =$$
$$\left(\vec{\nabla}.\vec{E} + \vec{\nabla}.\vec{B} + \dot{g},\ -\dot{\vec{E}} + \vec{\nabla} \times \vec{B} + \dot{\vec{B}} + \vec{\nabla} \times \vec{E} + \vec{\nabla} g\right)$$

Unified Field Equations

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) =$$

$$= \left(\ddot{\phi} - \vec{\nabla}.\dot{\vec{A}} + \vec{\nabla}.\dot{\vec{A}} + \vec{\nabla}.\vec{\nabla}\phi - \vec{\nabla}.\vec{\nabla} \times \vec{A},\ -\ddot{\vec{A}} - \vec{\nabla}\dot{\phi} + \right.$$
$$\left. \vec{\nabla} \times \dot{\vec{A}} + \vec{\nabla}\dot{\phi} - \vec{\nabla}\vec{\nabla}.\dot{\vec{A}} - \vec{\nabla} \times \dot{\vec{A}} - \vec{\nabla} \times \vec{\nabla}\phi + \vec{\nabla} \times \vec{\nabla} \times \vec{A}\right) =$$

$$= \left(\ddot{\phi} + \vec{\nabla}^2\phi,\ -\ddot{\vec{A}} - \vec{\nabla}^2\vec{A}\right) = 4\pi(\rho_u,\ \vec{J}_u)$$

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) =$$

$$= \left(\ddot{\phi} - \vec{\nabla}.\dot{\vec{A}} - \vec{\nabla}.\dot{\vec{A}} - \vec{\nabla}.\vec{\nabla}\phi + \vec{\nabla}.\vec{\nabla}\times\vec{A}, -\ddot{\vec{A}} - \vec{\nabla}\dot{\phi} + \vec{\nabla}\times\dot{\vec{A}} - \vec{\nabla}\dot{\phi} + \vec{\nabla}\vec{\nabla}.\dot{\vec{A}} + \vec{\nabla}\times\dot{\vec{A}} + \vec{\nabla}\times\vec{\nabla}\phi - \vec{\nabla}\times\vec{\nabla}\times\vec{A}\right) =$$

$$= \left(\ddot{\phi} - \vec{\nabla}^2\phi - 2\vec{\nabla}.\dot{\vec{A}}, -\ddot{\vec{A}} + \vec{\nabla}^2\dot{\vec{A}} - 2\vec{\nabla}\dot{\phi} + 2\vec{\nabla}\times\dot{\vec{A}}\right) =$$

$$= 4\pi(\rho_u, \vec{J}_u)$$

Conservation Laws

$$\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)(g, \vec{E} + \vec{B}) =$$

$$= \left(\ddot{g} + \vec{\nabla}^2 g, \ddot{\vec{E}} + \vec{\nabla}\times\dot{\vec{B}} + \vec{\nabla}\vec{\nabla}.\vec{E} - \vec{\nabla}\times\dot{\vec{E}} - \vec{\nabla}\times\vec{\nabla}\times\vec{B}\right) =$$

$$= \left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)4\pi(\rho_g + \rho_e, \vec{J}_g + \vec{J}_e) =$$

$$= 4\pi\left(\dot{\rho}_g + \dot{\rho}_e + \vec{\nabla}.\vec{J}_g + \vec{\nabla}.\vec{J}_e, \dot{\vec{J}}_g + \dot{\vec{J}}_e - \vec{\nabla}\rho_g - \vec{\nabla}\rho_e - \vec{\nabla}\times\vec{J}_g - \vec{\nabla}\times\vec{J}_e\right)$$

$$\dot{\rho}_e + \vec{\nabla}.\vec{J}_e = 0$$

$$\dot{\vec{J}}_g - \vec{\nabla}\rho_g - \vec{\nabla}\times\vec{J}_g = 0$$

If the differential operator acts on the hyperbolic equation, analogous results are obtained:

$$\left(\frac{\partial}{\partial t}, \vec{\nabla}\right)\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(g, \vec{E} + \vec{B}) =$$

$$= \left(\ddot{g} + \vec{\nabla}^2 g, \ddot{\vec{E}} + \ddot{\vec{B}} + \vec{\nabla}\vec{\nabla}.\vec{E} - \vec{\nabla}\times\vec{\nabla}\times\vec{B} - \vec{\nabla}\times\vec{\nabla}\times\vec{E}\right) =$$

$$= \left(\frac{\partial}{\partial t}, \vec{\nabla}\right)4\pi(\rho_g + \rho_e, \vec{J}_g + \vec{J}_e) =$$

$$= 4\pi\left(\dot{\rho}_g + \dot{\rho}_e - \vec{\nabla}.\vec{J}_g - \vec{\nabla}.\vec{J}_e, \dot{\vec{J}}_g + \dot{\vec{J}}_e + \vec{\nabla}\rho_g + \vec{\nabla}\rho_e + \vec{\nabla}\times\vec{J}_g + \vec{\nabla}\times\vec{J}_e\right)$$

There are two conservation laws here, charge conservation for electromagnetism in the scalar, and a vector conservation for gravity.

$$\dot{\rho}_e - \vec{\nabla}.\vec{J}_e = 0$$

$$\dot{\vec{J}}_g + \vec{\nabla}\rho_g + \vec{\nabla} \times \vec{J}_g = 0$$

Gauge Transformations

$$(\phi, \vec{A}) \longrightarrow (\phi', \vec{A}') = \left(\phi - \dot{\lambda} - \vec{\nabla}.\vec{\Lambda}, \vec{A} + \vec{\nabla}\lambda - \dot{\vec{\Lambda}} + \vec{\nabla} \times \vec{\Lambda}\right) =$$

$$(\phi, \vec{A}) + \left(-\frac{\partial}{\partial t}, \vec{\nabla}\right)(\lambda, \vec{\Lambda})$$

Equations of Motion

$$\left(\gamma, \gamma\vec{\beta}\right)(g, \vec{E} + \vec{B}) =$$

$$=$$

$$\left(\gamma g - \gamma\vec{\beta}.\vec{E} - \gamma\vec{\beta}.\vec{B}, \gamma\vec{E} + \gamma\vec{\beta} \times \vec{B} + \gamma\vec{B} + \gamma\vec{\beta} \times \vec{E} + \gamma\vec{\beta} g\right) =$$

$$= \left(\frac{\dot{W}}{m} + \frac{\dot{W}}{e}, \frac{\dot{\vec{P}}}{m} + \frac{\dot{\vec{P}}}{e}\right)$$

■ Unified Equations of Motion

Repeat the exercise from above, but this time, look to the potentials.

$$\left(\gamma, \gamma\vec{\beta}\right)\left(\frac{\partial}{\partial t}, -\vec{\nabla}\right)(\phi, -\vec{A}) =$$

$$\left(\gamma, \gamma\vec{\beta}\right)\left(\dot{\phi} - \vec{\nabla}.\vec{A}, -\dot{\vec{A}} - \vec{\nabla}\phi + \vec{\nabla} \times \vec{A}\right) =$$

$$= \left(\gamma\dot{\phi} - \gamma\vec{\nabla}.\dot{\vec{A}} + \gamma\vec{\beta}.\dot{\vec{A}} + \gamma\vec{\beta}.\vec{\nabla}\phi - \gamma\vec{\beta}.\vec{\nabla} \times \vec{A},\right.$$
$$-\gamma\ddot{\vec{A}} - \gamma\vec{\nabla}\dot{\phi} + \gamma\vec{\nabla} \times \dot{\vec{A}} + \dot{\phi}\gamma\vec{\beta} -$$
$$\left.\vec{\nabla}.\vec{A}\gamma\vec{\beta} - \gamma\vec{\beta} \times \dot{\vec{A}} - \gamma\vec{\beta} \times \vec{\nabla}\phi + \gamma\vec{\beta} \times \vec{\nabla} \times \vec{A}\right)$$

That is pretty complicated! The key to simplifying this equation is to see what happens for light, where dt/dx = dx/dt. Gamma blows up, but if the equation is over gamma, that problem becomes a scaling factor. With beta equal to one, a number of terms cancel, which can be seen more clearly if the terms are written out explicitly.

$$= \left(\dot{\phi} - \frac{\partial}{\partial \vec{X}} \cdot \frac{\partial \vec{A}}{\partial t} + \frac{\partial \vec{X}}{\partial t} \cdot \frac{\partial \vec{A}}{\partial t} + \frac{\partial \vec{X}}{\partial t} \cdot \frac{\partial}{\partial \vec{X}} \phi - \frac{\partial \vec{X}}{\partial t} \cdot \frac{\partial}{\partial \vec{X}} \times \vec{A},\right.$$
$$- \dot{\vec{A}} - \frac{\partial}{\partial \vec{X}} \frac{\partial \phi}{\partial t} + \frac{\partial}{\partial \vec{X}} \times \frac{\partial \vec{A}}{\partial t} + \frac{\partial \phi}{\partial t} \frac{\partial \vec{X}}{\partial t} -$$
$$\left.\frac{\partial}{\partial \vec{X}} . \vec{A} \frac{\partial \vec{X}}{\partial t} - \frac{\partial}{\partial \vec{X}} \times \frac{\partial \vec{A}}{\partial t} - \frac{\partial}{\partial \vec{X}} \times \frac{\partial}{\partial \vec{X}} \phi + \frac{\partial}{\partial \vec{X}} \times \frac{\partial}{\partial \vec{X}} \times \vec{A}\right)$$

It would take a real mathematician to state the proper constraints on the three pairs of cancellations that happen when velocities get flipped. There are also a pair of vector identities, presuming simple connectedness. This leads to the following equation:

$$= \left(2 \dot{\phi}, - \dot{\vec{A}} - \frac{\partial}{\partial \vec{X}} . \vec{A} \frac{\partial \vec{X}}{\partial t} + \frac{\partial}{\partial \vec{X}} \times \frac{\partial}{\partial \vec{X}} \times \vec{A}\right)$$

The scalar change in energy depends only on the scalar potential, and the 3−vector change in momentum only depends on the 3−vector A.

Strings

$$dq^2 = \left(da_0^2 e_0^2 + da_1^2 \frac{e_1^2}{9} + da_2^2 \frac{e_2^2}{9} + da_3^2 \frac{e_3^2}{9},\right.$$
$$\left.2 da_0 da_1 e_0 \frac{e_1}{3}, 2 da_0 da_2 e_0 \frac{e_2}{3}, 2 da_0 da_3 e_0 \frac{e_3}{3}\right) =$$
$$= (\text{interval}^2, 3 - \text{string})$$

Dimensionless Strings

$$dq^2 = \frac{c^5}{G h} \left(da_0^2 e_0^2 + da_1^2 \frac{e_1^2}{9 c^2} + da_2^2 \frac{e_2^2}{9 c^2} + da_3^2 \frac{e_3^2}{9 c^2},\right.$$
$$\left.2 da_0 da_1 e_0 \frac{e_1}{3 c}, 2 da_0 da_2 e_0 \frac{e_2}{3 c}, 2 da_0 da_3 e_0 \frac{e_3}{3 c}\right)$$

As far as the units are concerned, this is relativistic (c) quantum (h) gravity (G). Take this constants to zero or infinity, and the difference of a quaternion blows up or disappears.

Behaving Like a Relativistic Quantum Gravity Theory

Case 1: Constant Intervals and Strings

$$T : dq \to dq' \text{ such that scalar } (dq^2) = \text{scalar } (dq'^2) \text{ and vector } (dq^2) = \text{vector } (dq'^2)$$

Case 2: Constant Intervals

$$T : dq \to dq' \text{ such that scalar}(dq^2) = \text{scalar}(dq'^2) \text{ and vector}(dq^2) \neq \text{vector}(dq'^2)$$

Case 3: Constant Strings

$$T : dq \to dq' \text{ such that scalar}(dq^2) \neq \text{scalar}(dq'^2) \text{ and vector}(dq^2) = \text{vector}(dq'^2)$$

Case 4: No Constants

$$T : dq \to dq' \text{ such that scalar}(dq^2) \neq \text{scalar}(dq'^2) \text{ and vector}(dq^2) \neq \text{vector}(dq'^2)$$

In this proposal, changes in the reference frame of an inertial observer are logically independent from changing the mass density. The two effects can be measured separately. The change in the length−time of the string will involve the inertial reference frame, and the change in the interval will involve changes in the mass density.

Conclusions

■ What's been done

It is an old dream (or perhaps more accurately, a recurring nightmare) to express laws of physics using quaternions. In these web pages, quaternion operators were employed to express central laws of physics: Newton's second law, the Maxwell equations, and the Klein−Gordon equation for relativistic quantum mechanics. Applications of quaternions to special relativity were done in detail, with over 50 problems worked out explicitly. Quaternions do not make problem solving easy. Rather, they help unite the laws themselves. Significantly, an analysis of the length of a quaternion interval if the origin is moved establishes a connection to the machinery of general relativity, the affine parameter.

■ What's new

One might suspect that the reason for the success claimed above is that nothing proposed is new. After all, quaternions are a linear combination of tensors of rank zero and one, and while used in a new way here, does anything genuinely novel appear?

I believe that a new very powerful idea drives this work, namely, that events represented as quaternion are a topological algebraic field. This implies that any collection of events can be generated by an appropriate quaternion function. Scalars and vectors mix under multiplication, so quaternions are a mixed representation.

A new view of relativistic quantum mechanics was outlined. The Klein−Gordon equation is a scalar equation. When quaternion operators are employed, the Klein−Gordon equation is part of a larger set, including a scalar and vector identity analogous to the Maxwell equations. These additional identities are also valid by the conventional analysis, but they do not naturally arise, so the parallel to the Maxwell equations is less clear.

A new link to general relativity has been proposed which is slightly different. The invariant interval in special relativity was the first term of the difference between two events quaternions squared. If the origin changes, then the first term of the difference between the two event quaternions and the origin quaternions squared is similar to the square of the affine parameter of general relativity. The only difference lies in the cross term.

Every event, every function, every operator used was a member of the field of quaternions. This might strike some as a comic reliance on a solitary tool. I prefer to think of it as a great democratic principle. Physics is impressively democratic, with each photon or electron obeying the same collection of laws interchangeably. The mathematics underlying the laws of physics should reflect this interchangeability.

■ What needs to be done

Nothing presented was proven rigorously. There were few references to the literature. This body of work is more like a skeleton of work, a reflection of the author's semi−formal training and isolation from the professional physics community. There is a need to flesh these ideas out only if there is the potential for new insights. Although I appreciate the standard approach, I feel like I have gained new insight into why Maxwell's equations are necessary, have a new way to view relativistic quantum mechanics, and cling to a novel toehold on general relativity. And since each of these is a quaternion, it becomes possible to mix and match them to create new areas of study. I hope this work generates interest in the physics community.

More problems need to be solved. My upcoming focus will be the Dirac function and Fourier analysis using quaternions. If I can build these functions, it should be possible to approach problems in electromagnetism and quantum mechanics. I won't be easy, it never is, but it might be elegant. And maybe I'll dabble a little with curves...