Gravity

doug sweetser@alum.mit.edu

Gravity by Analogy to EM

Abstract

A Lagrange density for gravity is proposed based on a strict analogy to the classical Lagrangian for electromagnetism. For local covariant coordinates where the connection is zero, the field equations are a four−dimensionalwave equation. The classic field equations contain both the Maxwell equations and Newton's field equations under certain conditions. The four−dimensionalwave equation has been quantized before. The scalar and longitudinal modes of emission are interpreted as 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−velocitysolution 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−orderdifferential 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−Newtonianaccuracy. 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 for an electrically charged hollow sphere. 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 gravitational field is zero. The field equations have been quantized before, but new interpretations will flow from the unification effort. A link to the 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−orderparameterized post−Newtonianaccuracy. 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:

$$\mathcal{L}_{EM} = -\frac{m}{\gamma V} - \frac{q}{c^2 V}\frac{U^\mu}{\gamma} A_\mu - \frac{1}{4 c^2}\left(A^{\mu,\nu} - A^{\nu,\mu}\right)\left(A_{\mu,\nu} - A_{\nu,\mu}\right)$$

An analogous Lagrangian for gravity would also contain these three components, but three changes are required. First, gravity depends on mass, not 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. Therefore 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 standard derivative (symbolized by a comma) with a covariant derivative (symbolized by a semicolon):

$$\mathcal{L}_g = -\frac{m}{\gamma V} + \frac{\sqrt{G}\, m}{c^2 V}\frac{U^\mu}{\gamma} A_\mu - \frac{1}{4 c^2}\left(A^{\mu;\nu} + A^{\nu;\mu}\right)\left(A_{\mu;\nu} + A_{\nu;\mu}\right)$$

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:

$$\mathcal{L}_{gEM} = -\frac{m}{\gamma V} - \frac{q - \sqrt{G}\, m}{c^2 V}\frac{U^\mu}{\gamma} A_\mu - \frac{1}{2 c^2} A^{\mu;\nu} A_{\nu;\mu}$$

$$= -\frac{m}{\gamma V} - \frac{q - \sqrt{G}\, m}{c^2 V} (\phi - \vec{A} \cdot \vec{v}) - \frac{1}{2 c^2} A^{\mu,\nu} A_{\nu,\mu} - \frac{1}{2 c^2} \Gamma^{\omega}{}_{\mu\nu} (\Gamma_{\rho}{}^{\mu\nu} A^{\rho} A_{\omega} - 2 A^{\mu,\nu} A_{\omega})$$

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:

$$\mathcal{L}_{gEM} = -\frac{m}{V} \sqrt{1 - \left(\frac{\partial \vec{R}}{\partial t}\right)^2} - \frac{q - G^{.5}\, m}{c^2 V} (\phi - \vec{A} \cdot \vec{v}) - \frac{1}{2 c^2} \left(\left(\frac{\partial \phi}{\partial t}\right)^2 - (\vec{\nabla} \phi)^2 - \left(\frac{\partial \vec{A}}{\partial t}\right)^2 + (\nabla \vec{A})^2 \right)$$

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.

■ Classical Field Equations

The field equations can be found by applying the Euler−Lagrangeequations to the Lagrange density (assuming the connection is zero for simplicity):

$$\Box^2 A^{\mu} = \frac{q - \sqrt{G}\, m}{V} \frac{U^{\mu}}{\gamma c}$$

The fields are expressed in terms of the potential. The symmetric and antisymmetric field strength tensors are very simpliar, differeing 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:

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} - c \vec{\nabla} \phi$$

$$\vec{e} = \frac{\partial \vec{A}}{\partial t} - c \vec{\nabla} \phi$$

$$\vec{B} = c \left(\frac{\partial}{\partial y} - \frac{\partial}{\partial z}, \frac{\partial}{\partial x} - \frac{\partial}{\partial z}, \frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right) \vec{A} = c \vec{\nabla} \times \vec{A}$$

$$\vec{b} = c \left(- \frac{\partial}{\partial y} - \frac{\partial}{\partial z}, - \frac{\partial}{\partial x} - \frac{\partial}{\partial z}, - \frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right) \vec{A} \equiv \vec{\nabla} \times \vec{A}$$

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:

$$g = \left(\frac{\partial \phi}{\partial t}, -c \frac{\partial A_x}{\partial x}, -c \frac{\partial A_y}{\partial y}, -c \frac{\partial A_z}{\partial z} \right) = \partial^{\mu} A^{\mu}$$

The diagonal of the field strength tensor A^mu ;nu 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:

$$\rho_q - \rho_m = \frac{\partial^2 \phi}{\partial t^2} - c^2 \vec{\nabla}^2 \phi$$

$$= \frac{c}{2} (\vec{\nabla} \cdot \vec{E} + \vec{\nabla} \cdot \vec{e}) + \frac{\partial g^0}{\partial t}$$

This eqation 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:

$$\rho_m = c^2 \vec{\Box}^2 \phi$$

$$\text{iff } \frac{\partial \vec{A}}{\partial t} = -c \vec{\nabla} \phi$$

This equation should be consistent with special relativity without modification. The classical Newtonian field equation arises from these physical constraints:

$$\rho_m = c^2 \vec{\nabla}^2 \phi$$

$$\text{iff } \frac{\partial \vec{A}}{\partial t} = -c \vec{\nabla} \phi \text{ and } \frac{\partial g^0}{\partial t} = 0$$

Every aspect of classical Newtonian gravity can be represented by this proposal under these constraints.

Gauss' law appears under the following conditions:

$$\rho_q = \frac{\partial^2 \phi}{\partial t^2} - c^2 \vec{\nabla}^2 \phi$$

$$\text{iff } \frac{\partial \vec{A}}{\partial t} = c \vec{\nabla} \phi$$

Repeat the exercise for the vector equation.

$$-\vec{J}_m + \vec{J}_q = \frac{\partial^2 A}{\partial t^2} - c^2 \nabla^2 \vec{A}$$

$$= \frac{1}{2}\left(-\frac{\partial \vec{E}}{\partial t} + c\,\vec{\nabla} \times \vec{B} + \frac{\partial \vec{e}}{\partial t} + c\,\vec{\nabla} \underline{\times}\, \vec{b}\right) + c\,\vec{\nabla} g^u$$

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 [x^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 [x^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−rankfield 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−vectorsand a curved metric. Start with the standard model Lagrangian:

$$\mathcal{L}_{SM} = \bar{\Psi}\,\gamma^{\mu}\,D_{\mu}\,\Psi$$

where

$$D_{\mu} = \partial_{\mu} - ig_{EM}\,YA_{\mu} - ig_{weak}\,\frac{\tau^{a}}{2}\,W^{a}_{\mu} - ig_{strong}\,\frac{\lambda^{b}}{2}\,G^{b}_{\mu}$$

The electromagnetic potential A_mu is a complex−valued4−vector.The only way to form a scalar with a 4−vectoris 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−vectoris 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−symmetricfield strength tensor:

$$F^{\mu}_{EM} = q\,\frac{U_{\nu}}{c}\,(A^{\mu,\nu} - A^{\nu,\mu})$$

Form an analogous force for gravity using the same substitutions as before:

$$F^{\mu}_{g} = -\sqrt{G}\,m\,\frac{U_{\nu}}{c}\,(A^{\mu;\nu} + A^{\nu;\mu})$$

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:

$$F^{\mu}_{gEM} = \left(q - \sqrt{G}\ m\right) \frac{U_{\nu}}{c} A^{\mu;\nu} - \left(q + \sqrt{G}\ m\right) \frac{U_{\nu}}{c} A^{\nu;\mu}$$

If q >> G^.5 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−handside 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:

$$0 = \left(q - \sqrt{G}\ m\right) \frac{\partial x_{\nu}}{c\, \partial \tau} A^{\mu,\nu} - \left(q + \sqrt{G}\ m\right) \frac{\partial x_{\nu}}{c\, \partial \tau} A^{\nu,\mu} - \frac{2}{c} m \Gamma_{\omega}{}^{\mu\nu} U_{\nu} U^{\omega}$$

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.

Since the goal in this section is to study gravity, not electromagnetism, work with a potential that can only contribute to gravity, not electromagnetism. Since the first component of the potential does not appear in the antisymmetric field strength tensor, work with a potential with the form: A_mu = (A_0, 0).

The next task is to find a solution to the unified field equations. 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/sigm^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:

$$\frac{\partial \phi}{\partial t} = \frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\,(k^2)$$

$$c\,\frac{\partial A_x}{\partial x} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\,(k^2)$$

$$c\,\frac{\partial A_y}{\partial y} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\,(k^2)$$

$$c\,\frac{\partial A_z}{\partial z} = -\frac{c^2\,k}{\sqrt{G}\,\sigma^2} + O\,(k^2)$$

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 appliex exclusively to gravity is saught. The sign of the spring constant k does not effect the solution to the 4D wave field equations. The sign of the spring constant k does change the derivative of the potential. Therefore a potential that only has derivatives along the diagonal of the field strength tensor can be constructed from two potentials that differ by string constants that either constructively interfere to create a non−zeroderivative, or destructively interfere to eliminate a derivative.

diagonal SHO A^{μ} =

$$\frac{c}{\sqrt{G}}\left(1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}t\right)^2\right)+1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}t\right)^2\right),\right.$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}t\right)^2\right)+1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}t\right)^2\right),$$

$$1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}t\right)^2\right)+1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}t\right)^2\right),$$

$$\left.1\Big/\left(\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}t\right)^2\right)+1\Big/\left(\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}x\right)^2+\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}y\right)^2+\left(\frac{1}{\sqrt{2}}+\frac{k}{\sigma^2}z\right)^2-\left(\frac{1}{\sqrt{2}}-\frac{k}{\sigma^2}t\right)^2\right)\right)$$

Take the contravariant derivative of this potential, keeping only the terms to first order in the spring constant k. The contravariant derivative flips the sign for the three−vector.

$$A^{\mu,\nu} = \frac{c^2}{\sqrt{G}} \begin{pmatrix} \frac{k}{\sigma^2} & 0 & 0 & 0 \\ 0 & \frac{k}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{k}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{k}{\sigma^2} \end{pmatrix}$$

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 −t^2 =R’^2. In this way 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:

$$\left(\frac{\partial t}{\partial \tau}, \frac{\partial \vec{R}}{c\, \partial \tau} \right) \longrightarrow \left(\frac{\partial t}{\partial\, |R|}, \frac{\partial \vec{R}}{c\, \partial\, |R|} \right) = (0, \hat{R})$$

Plug these three assumptions into force equation:

$$F_g^\mu = \left(0, -\frac{GMm}{R^2} \hat{R}\right)$$

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:

$$m \frac{\partial^2 \vec{R}}{\partial t^2} + \frac{\partial \vec{R}}{\partial t} \frac{c\,\partial m}{\partial\,|R|} = -\frac{GMm}{R^2}\hat{R}$$

For a point source, the dm/dtau 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−orderdifferential equations. The equation can be simplified to a set of first−orderdifferential 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−t^2 = R’^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).

■ 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−vectorsrequires 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 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

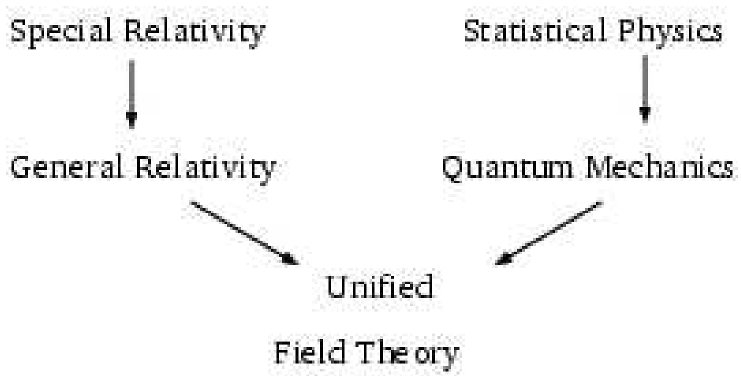
■ Abstract

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−dimensionaltopological 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−matrixfield strength quaternion for electricity and an odd, antisymmetric−matrixfield 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−Bohmeffect, 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 longstanding 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 some 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−dimensionalgeometry is dictated by insights garnered from physics. Events in spacetime are composed of a scalar for time and a 3−vectorfor space. The four−dimensionaltopological 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−independentway. 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−independenceexplicit. In special relativity, regions in spacetime are delimited by the light cone, where the net change in 3−spaceis equal to the net change in time. The parity between changes in 3−spaceand 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−orderRiemannian quaternion differential equations create force fields for gravity, electricity, and magnetism. Second−orderdifferential equations create dynamic field equations for gravity, the Maxwell equations for electromagnetism, and a classical counterpart to the Aharanov−Bohmeffect of quantum mechanics. Third−orderdifferential 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−basiscomposed 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−independent4−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−dimensionalalgebraic 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−spaceover the change in time is equal to one. Physics therefore indicates parity between the total 3−vectorand the scalar, instead of weighing all four equally. A coordinate−independentRiemannian 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 scaling factor of a third for the 3−vectorplays a vital role in the definition of a regular function in the third paper of this series.

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:

$$-\frac{\hat{i}_n^{\,2}}{9}\frac{\partial E_n}{\partial i_n}-\frac{\hat{i}_n E_n}{9}\frac{\partial \hat{i}_n}{\partial i_n}=4\pi\rho,\quad n=1,2,3$$

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−zeroenergy 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−basisvectors, 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−vectorfor each quaternion playing an analogous role. The difference is that unit 3−vectorsdo not all have to point in the same direction. Based on the angle between them, two different unit 3−vectorshave 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−commutingalgebras 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−spacelocation and will have a non−zerolength if it ends up at a different location and time. With quaternion products, the 3−ropeis 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−ropeis covariant, because it is known how it changes. Given the utility of duality, 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−ropeis 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−spacebasis vectors are inversely related, the magnitude of the product of the time basis vector with each 3−spacebasis vector will be constant even if the basis vectors themselves are dynamic. This hypothesis asserts the existence of 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−ropeis 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−forceis 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−valuedquaternions 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−zeroquaternion could equal zero if the values of t, x, y, and z were complex. This paper involves the constraint of working exclusively with 4−dimensionalalgebraic fields. Therefore, no matter how salutary the work with complex−valuedquaternions, 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−dimensionalcontinuum. 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−dimensionalcontinuum" 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−diagonaland 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−scalarformed from a differential quaternion acting on a potential. To be invariant under an arbitrary gauge transformation, the quaternion−scalarmust 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−connectedtopology). 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−matrixforce field would be linked to the electromagnetic source equation, Ampere’s law.

Generate the other two Maxwell equations:

$$\frac{-(\Box \mathrm{Vector}(\Box^* A^*) + \Box^* \mathrm{Vector}(\Box A))^*}{2} =$$

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

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

$$= (4\pi\rho,\ \vec{0}).$$

This is Gauss’ and Faraday’s law. Again, if the vector operator had not been used, the time derivative of the symmetric−matrixforce 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−vectorapproach.

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−invariantway. 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−vectoracts on a 4−vectorpotential in such a way as to create an antisymmetric second−ranktensor. The unified field hypothesis proposed involves a quaternion differential operator acting on a quaternion potential:

$$\Box^* A^* = \left(\frac{\partial \phi}{\partial t} - \vec{\nabla}.\vec{A}, \; -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}\phi + \vec{\nabla} \times \vec{A} \right)$$

This is a natural suggestion with this algebra. The antisymmetric−matrixcomponent 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−vectorfunction 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:

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

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

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

$$\left.+2\vec{\nabla}\times\frac{\partial \vec{A}}{\partial t} + \vec{\nabla}\times\vec{\nabla}\phi\right).$$

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−vectorcurrent density J? Now there is a current that transforms like a pseudo−currentdensity. The inclusion of a pseudo−currentalong with a current making the proposal more complete. The volume integral of this pseudo−currentdensity 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−vectorcurrent composed of the difference between the time derivative of the magnetic field and the curl of the electric field. The Aharonov−Bohmeffect (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−currentdensity is quite unusual, transforming differently under space inversion than the electric current density. One might imagine that a Lorentz transformation would shift this pseudo−currentdensity into a pseudo−chargedensity. This does not happen however, because the vector identity involving the divergence of a curl still applies. The Aharonov−Bohmphenomenon, 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−matrixgravitational 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:

$$J = J_m \text{ iff } \vec{E} = \vec{B} = \vec{0}$$

$$J = J_e + \vec{J}_{AB} \text{ iff } g = 0.$$

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−Bohmpseudo−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:

$$4\pi J^* = \Box\Box^* A^* =$$

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

$$= \left(\frac{\partial g}{\partial t} - \vec{\nabla}.\vec{E} - \vec{\nabla}.\vec{B},\right.$$

$$\left. \vec{\nabla} g + \frac{\partial \vec{E}}{\partial t} + \vec{\nabla} \times \vec{B} + \frac{\partial \vec{B}}{\partial t} + \vec{\nabla} \times \vec{E} \right).$$

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−currentmust be included. This may be the differential form of a classical Aharonov−Bohmeffect.

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−zerosolutions 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:

$$\Box\Box^{*}\left(\frac{1}{t^2-x^2-y^2-z^2},\vec{0}\right)=\left(\frac{4\,(3\,t^2+x^2+y^2+z^2)}{(t^2-x^2-y^2-z^2)^3},\vec{0}\right)$$

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−zeropart 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 this paper. 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.

$$\text{Scalar}(\Box J_e^{*}) = \frac{\partial\rho_e}{\partial t} + \vec{\nabla}\cdot\vec{J}_e = 0$$

$$\text{Scalar}(\Box\vec{J}_{AB}^{\,*}) = \nabla\cdot\vec{J}_{AB}^{\,*} = 0$$

$$\text{Vector}(\Box J_m^{*}) = -\frac{\partial\vec{J}_m}{\partial t} + \vec{\nabla}\rho_m - \vec{\nabla}\times\vec{J}_m = \vec{0}$$

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−Bohmcurrent, 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−vectorcovariant, 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:

$$(\Box'^{*}A'^{*})_{\text{Residual}} = \left(0,\ (\gamma^2\beta^2 - 1)\frac{\partial A_x}{\partial t} + (\gamma^2 - 1)\frac{\partial\phi}{\partial x},\right.$$
$$\left. -2\gamma\beta\left(\frac{\partial A_z}{\partial t} + \frac{\partial A_y}{\partial x}\right),\ 2\gamma\beta\left(\frac{\partial A_y}{\partial t} - \frac{\partial A_z}{\partial x}\right)\right).$$

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−dependentphase factor. If so, it might provide a test for the theory.

The mechanics of the Lorentz transformation itself might require careful re−examinationwhen 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−vectorcan 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−vectorwith 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:

$$A \to A' = A - \Box^* \lambda .$$

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−vectorfield:

$$A \to A' = A - \Box^* \vec{\Lambda} .$$

This time the gravitational force field is invariant under a 3−vectorgauge field transformation. Additional constraints can be placed on the 3−vectorgauge field to preserve a chosen electromagnetic invariant. For example, if the difference between the two electromagnetic fields is to remain invariant, then the 3−vectorgauge field must be the solution to an elliptical equation. Other classes of invariants could be examined.

The scalar and 3−vectorgauge 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:

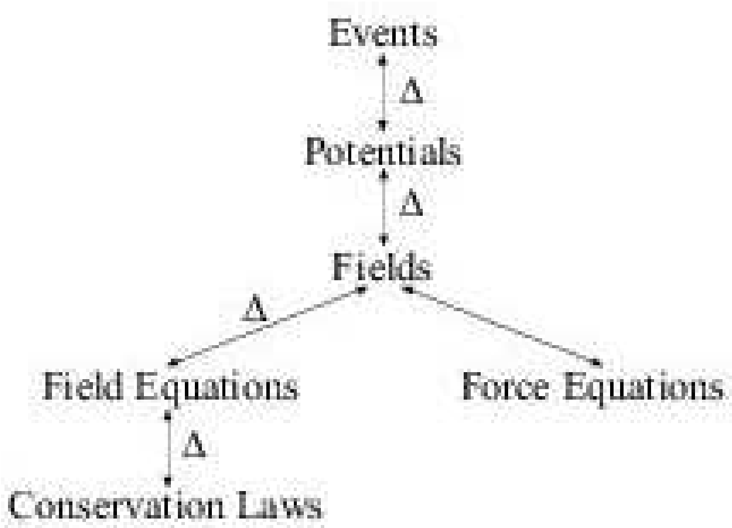
$$A \to A' = A - \Box^* \Lambda .$$

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−matrixforce 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−Bohmeffect. 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. To work within the guidelines of this paper, one should avoid solutions represented by complex−valuedquaternions.

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 photons in a vacuum, but that does not have to be the case for a nonhomogeneous equation. 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(−(JA*) −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, (dtau)^2 = e^−(2GM/c^2tau) dt^2 −e^(2GM/c^2 tau) dR^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−velocitysolution exists for the gravitational force equation for a system with an exponentially−decayingmass distribution. The dark matter hypothesis is not needed to explain the constant−velocityprofiles 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−potentialA^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−vectorfield 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 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:

$$\Box^2 A = 4 \pi \left(k J_{charge} + G J_{mass} \right)$$

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−invariantinterval squared. Like mass, the 4−potentialwill 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:

$$\frac{\Box^2 A}{|A|} = 4\pi\,(kJ_{charge} + GJ_{mass})$$

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:

$$t \;->\; t' = A + \frac{GM}{2\,c^2\,A}\,t$$

$$\vec{R} \;->\; \vec{R}' = \vec{B} + \frac{GM}{2\,c^2\,|\vec{B}|}\,\vec{R}$$

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 infusers 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:

$$L = -J^\mu A_\mu - \frac{1}{2}\left(\partial^\mu A_\mu\right)^2 - \frac{1}{4}\left(\partial^\mu A^\nu - \partial^\nu A^\mu\right)\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)$$

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:

$$L = \text{scalar}\left(-JA^* - \frac{1}{2}\Box^* A\,\Box A^*\right)$$

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−forceto a metric

A relativistic 4−forceis 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−forceof 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:

$$\left(\frac{\partial^2 t}{\partial\tau^2} + \frac{GM}{c^2\,\tau^2}\frac{\partial t}{\partial\tau},\ \frac{\partial^2\vec{R}}{\partial\tau^2} - \frac{GM}{c^2\,\tau^2}\frac{\partial\vec{R}}{\partial\tau}\right) = (0,\ \vec{0})$$

Solve these second−orderdifferential equations for the spacetime position:

$$t = c_1 \left(\tau e^{\frac{GM}{c^2 \tau}} - \frac{GM}{c^2} \operatorname{Ei}\left(\frac{GM}{c^2 \tau} \right) \right) + C_2$$

$$\vec{R} = \vec{C}_1 \left(\tau e^{-\frac{GM}{c^2 \tau}} + \frac{GM}{c^2} \operatorname{Ei}\left(-\frac{GM}{c^2 \tau} \right) \right) + \vec{C}_2$$

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 spacetime 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−orderin M/|tau|:

$$\partial \tau^2 = \left(1 - 2 \frac{G M}{c^2 \tau} + 2 \left(\frac{G M}{c^2 \tau}\right)^2\right) \partial t^2 -$$

$$- \left(1 + 2 \frac{G M}{c^2 \tau} + 2 \left(\frac{G M}{c^2 \tau}\right)^2\right) \partial R^2 + O\left(\left(\frac{G M}{c^2 \tau}\right)^3\right)$$

Contrast this with the Schwarzschild solution in isotropic coordinates expanded to second order in M/R (MTW, eq. 31.22):

$$\partial \tau^2 = \left(1 - 2 \frac{G M}{c^2 R} + 2 \left(\frac{G M}{c^2 R}\right)^2\right) \partial t^2 -$$

$$- \left(1 + 2 \frac{G M}{c^2 R} + 2.5 \left(\frac{G M}{c^2 R}\right)^2\right) \partial R^2 + O\left(\left(\frac{G M}{c^2 R}\right)^3\right)$$

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−Newtonianaccuracy, 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−sourcemass 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:

$$\left(\gamma \frac{\partial m}{\partial \tau} + \frac{\gamma G M}{c^2 | \tau |^2} m, \gamma \vec{\beta} \frac{\partial m}{\partial \tau} - \frac{\gamma \vec{\beta} G M}{c^2 | \tau |^2} m\right) = (0, \vec{0})$$

Solve for the mass flow:

$$(\gamma m, \gamma \vec{\beta} m) = \left(c e^{\frac{G M}{c^2 |\tau|}}, \vec{C} e^{-\frac{G M}{c^2 |\tau|}}\right)$$

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−energydensity. 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−potentialA and the 4−basisI. 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−dimensionalautomorphic functions, we may have a deep appreciation of Nature's methods.

There is a physical explanation for gravity −it is a local, nonlinear, four−dimensionalsimple 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−stringwill 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/e1^2is 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.

$$\frac{1}{G} \to \frac{\text{mass time}^2}{\text{distance}^3} \qquad \frac{1}{h} \to \frac{\text{time}}{\text{mass distance}^2} \qquad c^5 \to \frac{\text{distance}^5}{\text{time}^5}$$

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−dimensionalquaternion, 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

$$T : dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) = \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) \mathrel{!=} \mathrm{vector}(dq'^2)$$

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

$$T : dq \to dq' \text{ such that } \mathrm{scalar}(dq^2) \mathrel{!=} \mathrm{scalar}(dq'^2) \text{ and } \mathrm{vector}(dq^2) = \mathrm{vector}(dq'^2)$$

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−inertialreference 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

$$T : dq \to dq' \text{ such that } \text{scalar}(dq^2) \neq \text{scalar}(dq'^2) \text{ and } \text{vector}(dq^2) \neq \text{vector}(dq'^2)$$

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.

■ 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 post, 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 post (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 in this post 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.

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

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−vectorwhich 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

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

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−valuedwave 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−−indeedthe selection of such a framework is an essential pre−requisitefor theoretical research−−providedthe 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−linearityinto quantum theory, which require a change in the formalism. It is easy to write non−linearquaternion functions. Near the end of this post 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−physicalway. 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 post 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−linearquaternion 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−classicalsense?

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−fieldof 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−spacequestion 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).

$$N(k) = (A + k B)^* (A + k B)$$

$$= A^* A + k (A^* B + B^* A) + k\hat{}2 B^* B$$

Find the extremum of the parameterized norm.

$$\frac{dN}{dk} = 0 = A^* B + B^* A + 2 k B^* B$$

The extremum is a minimum

$$\frac{d^2 N}{dk^2} = 2 B^* B >= 0$$

The minimum of a quaternion norm is zero. Plug the extremum back into the first equation.

$$0 <= A^* A - \frac{(A^* B + B^* A)^2}{2 B^* B} + \frac{(A^* B + B^* A)^2}{4 B^* B}$$

Rearrange.

$$(A^* B + B^* A)^2 <= 4 A^* A B^* B$$

Take the square root.

$$A^* B + B^* A <= 2 \sqrt{A^* A B^* B}$$

Add the norm of A and B to both sides.

$$A * A + A^* B + B^* A + B^* B <= A^* A + 2 \sqrt{A^* A B^* B} + B^* B$$

Factor.

$$N (A + B) = (A + B)^* (A + B) <= \left(\sqrt{A^* A} + \sqrt{B^* B}\right)^2$$

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−spaceto 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 post 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−linearshortest 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 notebook, 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 post 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 post.

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:

$$\mathrm{scalar}\left((dt,\ d\vec{X})^2\right) = dt^2 - d\vec{X}.d\vec{X}$$

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.

$$\mathrm{interval}^2 = \mathrm{scalar}\,(g\,dq\,g\,dq)$$

$$\text{if } g = (1, \vec{0}),$$

$$\text{then interval}^2 = dt^2 - d\vec{X}.d\vec{X}$$

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 :−)

$$\text{scalar}(((0, 1, 0, 0)(t, x, y, z))^2) =$$

$$= (-t^2 + x^2 - y^2 - z^2, \vec{0})$$

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:

$$\text{interval}^2 = \text{scalar}((g, \vec{G})(dt, d\vec{X})(g, \vec{G})(dt, d\vec{X})) =$$

$$= g^2 (dt^2 - d\vec{X}.d\vec{X}) - 4\, g\, dt\, \vec{G}.d\vec{X} +$$
$$(\vec{G}.d\vec{X})^2 - dt^2\, d\vec{G}.d\vec{G} - (\vec{G} x d\vec{X}).(\vec{G} x d\vec{X}) =$$

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−symmetricmetric. 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−Euclideannorm, 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 notebook, 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)