The breakdown of Euclid's Geometry in Linearized Relativity

Guy Vandegrift, Write State University Lake Campus, Celina, Ohio 45822

A linearized version of General Relativity is constructed by first postulating a simple field equation in the Lorentz Gauge. Equations of motion are obtained by assuming that particles are the limiting case of wavepackets of some linear wave equation. Seeking a dispersion relation, it is discovered that these waves travel in geodesics of a metric that was "perturbed" slightly to accommodate the gravitational redshift. Applying this linearized theory to the special case of a static gravitational field uncovers the fact that this perturbed metric violates Euclidean geometry.

I. A *Concept Map* and two "approaches" to General Relativity

The concept map of figure 1 depicts two approaches to the theory of General Relativity[1,2,3,4,5] This paper emphasizes a *"linearized approach"* that occupies the lower right-hand portion of the map. Though I hope readers will find this approach interesting, I hesitate to recommend it for an introductory course on General Relativity.[6] Such a course should instead focus on some version of the *"traditional approach"*, as depicted in the box labeled *Riemannian Calculus*. The student, confronted with this strange but useful calculus, faces a notation that seems alien to those who learned vector calculus through expressions such as, $\nabla^2\Phi$, $\nabla\times\mathbf{A}$, or $\nabla\cdot\mathbf{A}$. Furthermore, it can be disconcerting to realize that much of calculus had been learned by visualizing concepts under the false premise of Euclidean geometry. Nevertheless, any solid introductory course in General Relativity must face these challenges, because Riemannian calculus is essential to General Relativity.

Even this map's version of the *traditional approach* should not be construed as a literal course outline, because it is not necessary to derive equations of motion from the variational principle. Geodesics can be defined by setting the path length, $\int d\tau$, to an extremum so that the variation vanishes: $\delta\int d\tau=0$. Geodesics are the "straight lines" of Euclidean geometry, and the "great circles" on the surface of a sphere. The rank-2 metric tensor, $g_{\mu\nu}$, stipulates a generalized "distance" between two neighboring points[7] in space-time: $d\tau^2 = -\Delta x^\mu g_{\mu\nu}\Delta x^\nu$. But one can also use the directional derivative to obtain equations of motion. The result, $\ddot{x}^{\beta} = -\Gamma^{\beta}{}_{\mu\nu}\dot{x}^{\mu}\dot{x}^{\nu}$, is reminiscent of Newton's second law, $a^j \equiv \ddot{x}^j = -\partial_j V \equiv -\partial V/\partial x^j$. Here dots denote differentiation with respect to proper time, $\dot{x}^{\mu} \equiv dx^{\mu}/d\tau$. The motive for this emphasis on the variational principle will become clear when we explore the *linearized approach,* and point out that the variational principle is a direct consequence of the fact that particles can be viewed as wavepackets.[8,9]

As is customary in General Relativity, Latin letters are summed from 1 to 3, and Greek letters from 0 to 3. To keep the equations as familiar as possible, conventional and space-time variables, are treated as identical and interchangeable: $(t, \boldsymbol{r}) \equiv (t, x, y, z) \equiv (x^0, x^1, x^2\ x^3)$ Our units are such that, $\hbar = c = G$, where G is the gravitational constant. In these units, the static gravitational potential must be sufficiently small, $V<<1$, in order for the "weak" field limit commonly associated with linearized General Relativity to be valid. At a distance, r, outside a spherical mass, $V=-M/r$. Strictly speaking, $\hbar$ has no role in this theory, and may be taken as any small constant that causes wavepackets to act as particles.

With so many constants set equal to unity, we find that frequency-wavenumber equals the Hamiltonian-conjugate momentum of classical mechanics.[10] This can be turned to advantage because the concepts should be viewed as interchangeable. It is important that frequency-wavenumber, $K_\alpha=(-\omega, \boldsymbol{k}) \equiv (-\omega, k_1, k_2, k_3)$, be written in covariant (subscripted) form. The minus sign on the frequency (Hamiltonian) ensures that a plane wave takes on the form, $\exp(i\boldsymbol{k}\cdot\boldsymbol{\xi} - i\omega t) = \exp(iK_\alpha \xi^\alpha)$, where $\xi^\alpha = \Delta x^\alpha$ must be used, because in General Relativity, the space-time coordinate, x^α, is not a rank-1 tensor (4-vector), while the differential, Δx^α, is.

The geodesic postulate of the *traditional approach* is beautifully elegant, but also troubling: The fact that one *can* make a simple postulate with deep mathematical implications does not imply that one *must*. This paper makes no attempt to establish that Einstein's relativity is the only possible correct theory, or even the simplest theory that fits experimental evidence. But, perhaps some of the longing some students feel for "a simpler theory" can be alleviated, if the same result can be reached via an approach that strives to downplay the role of metric and intrinsic curvature.

Those not fluent with Riemannian calculus may find this *linearized approach* more accessible. Christoffel's connection coefficients are mentioned, but not required for the calculation, except to justify approximations. The covariant (subscripted) nature of frequency-wavenumber is established by analogy with a crystal lattice in solid state physics. The most complicated tensor that requires actual evaluation is rank-2 and diagonal. Even Hamilton's canonical equations of motion are provided with a simple derivation. As depicted in the map, the *linearized approach* fails to recover the advance of the perihelion of Mercury's orbit, which requires a higher order theory. But the *linearized approach* does capture the anomalous deflection of light by the Sun. And this can establish Euclidean geometry to be false in the physical universe.[11]

The box in the map labeled, $b = 1$, refers to a pesky parameter that equals unity if and only if the *linearized approach* recovers a well known version of Einstein's linearized theory. Most readers are invited to accept, $b=1$, based on experimental evidence concerning the deflection of light by the Sun's gravity. For the "purist", a theoretical argument is also presented in Appendix B.

The roles of *a priori* assumptions and *consequences* are largely reversed between the *traditional* and *linearized* approaches. This permits a fresh look at General Relativity that may help resolve unanswered questions. For example, what is it about gravity that requires such an exotic and alien metric? Here, the answer seems to involve the gravitational redshift of light. Both approaches (*linearized* and *traditional*) rely heavily on *Non-Relativistic Particle Motion,* as reflected by the two arrows flowing out of that box in the concept map. An essential ingredient of Newton's theory is that the path of a particle is independent of mass. The two approaches incorporate this fact, but in entirely different ways.

The *linearized approach* has two starting points. The first is labeled, *Simple Field Equation*: $\partial^{\alpha}\partial_{\alpha}\Phi^{\mu\nu} = -16\pi T^{\mu\nu}$, and relates a tensor field, $\Phi^{\mu\nu}$, to a source, $T^{\mu\nu}$. Although I "postulated" this field equation with *a priori* knowledge of its existence in linearized relativity, it is a natural and obvious "guess" given that similar field equations exist in *electrostatics*, and *electromagnetism*.[12] The factor -16π is an arbitrary constant, chosen to conform to conventional expressions in linearized relativity. Any non-zero constant would recover essentially the same theory. The other starting point for the *linearized approach* is the redshift of light, which must be constructed from physical arguments. (In contrast, the *traditional approach* derives this redshift as a consequence of the theory.)

The use of physical arguments in a theory as mathematically abstract as General Relativity is always problematical.[13,14] Fortunately, we have two physical arguments that independently yield the same formula for a shift in the frequency of light. Both arguments involve a static gravitational field, where the difference in gravitational potential obeys the weak field condition associated with linearized relativity, $\Delta V \ll 1$. One argument involves the mimicking of gravity by acceleration in an empty universe. Imagine a laser aimed at an observer while both are at rest. Now turn the laser on, while both observer and source simultaneously undergo identical accelerations along the axis defined by the laser and observer. Because of the time-delay between emission by the source and absorption by the observer, the observer will be travelling at a different speed than the laser was when it emitted the photon. Hence the observer will perceive a different frequency. At non-relativistic speeds, the time delay between emission and absorption is, $\Delta t = d/c = d$, where c is the speed of light and d is the distance between laser and observer. If the rate of acceleration is a, this time-delay causes the observers speed differ from that of the source by $\Delta v = a\Delta t = ad$. The non-relativistic Doppler shift in frequency for light is, $\Delta\omega/\omega = \pm v = \pm ad$, depending on the direction of the acceleration. Taking a to be the acceleration of gravity, the change in gravitational potential is therefore, $\Delta V = ad$. Thus, the mimicking of gravity by acceleration shifts frequency by $\omega\Delta V$. The same result can be obtained by energy conservation. The change in mass associated with photon absorption or emission obeys, $mc^2 = \hbar\omega$. In our units, this mass, the photon's frequency, and energy all equal to, ω. The Newtonian potential energy *("mgh")* for a photon is therefore ωV. Apparently the photon provides this energy by losing frequency. Hence, $\Delta\omega = -\omega V$.

This redshift raises great complications when one considers an important fact about linear wave equations: If the coefficients do not exhibit explicit time dependence, then solutions oscillate as, $\exp(-i\omega t)$, where ω is a constant of motion. Consider a static gravitational field, such as one might experience while standing on a planet. Take Schrödinger's equation, $i\partial_t\psi = -\frac{1}{2}\partial_i\partial^i\psi + V\psi$, as perhaps the most familiar example of a linear wave. If the potential is static, then $\partial_t V \equiv \partial V/\partial t = 0$, and solutions vary in time as, $\exp(-i\omega t)$. How can a wave exhibit a constant, ω, and still undergo a shift in frequency? The proposed solution to this conundrum involves will ultimately force the *linearized approach* to abandon Euclidean geometry.

In order to incorporate the gravitational redshift, it is postulated that t does not denote true time. It is tempting to view this as a conventional change of variables, from what one might call "actual time", t^*, to the timelike variable, $t = t(t^*, x^*, y^*, z^*)$. But this view is entirely wrong. Any so-called "actual" space and time variables associated with flat Minkowski space (t^*, x^*, y^*, z^*) are just as fictitious as are Cartesian variables (x^*, y^*) on the surface of a globe. A charming aspect of *this linearized approach* is that the nonexistence of global Minkowskian variables (t^*, x^*, y^*, z^*) need not be established at the beginning, provided one avoids using them. The reader is permitted to "pretend" that space-time is flat, as the theory is constructed. As is the case on a globe, Minkowskian coordinates are permitted within a small local neighborhood. Hence the differentials $(\Delta t^*, \Delta x^*, \Delta y^*, \Delta z^*)$ do exist, and will be used often.

Let ,Δt^*, be the perceived time as experienced by an observer in a static gravitational field, and let, $\Delta t \equiv \Delta x^0$, be the corresponding interval of the time-like variable, used to express the field and wave equations. As a stationary observer counts oscillations of this wave, the corresponding advance in phase obeys,

$$\omega^* \, \Delta t^* = \omega \, \Delta t \tag{1}$$

This will permit the perceived frequency, ω^*, to exhibit a redshit while ω remains a constant of motion as the photon rises against the gravitational field. As stated in the concept map, will result in a *"perturbation of the metric"*.

To summarize, by attempting to construct a theory of gravity in which the metric does **not** play a unique and remarkable role, the linearized approach recovers the weak-field limit of Einstein's theory. In doing so it makes a convincing case that gravity possesses some property that connects it so intimately with the metric. The step that seems to force this issue is the redshift of light. One can only imagine the havoc that would be wreaked upon *electrodynamics* , if the photon were discovered to have even the smallest amount of electric charge.

II. Why frequency-wavenumber is covariant (subscripted)

In a graduate-level course in plasma physics, circa 1980, professor Allan Kaufman wrote the following on the board, stating without proof that they could model the penetration of microwaves into a thermonuclear plasma:

$$\frac{dx^j}{dt} = \frac{\partial\omega}{\partial t}, \qquad \frac{dk_j}{dt} = -\frac{\partial\omega}{\partial x^j}, \qquad \frac{d\omega}{dt} = \frac{\partial\omega}{dt}. \tag{2}$$

The implication was that **all** linear waves exhibit wavepacket solutions that obey these Hamiltonian (or *canonical*) equations of motion, in the eikonal limit of sufficiently short wavelength and high frequency. Goldstein's text on *classical mechanics* establishes the converse: For any classical Hamiltonian, there is an associated linear wave equation[10]. Although I know of no proof, I strongly suspect that (2) holds for wavepackets of any linear wave, with few or no exceptions, in the eikonal limit.

General Relativity is one of many topics in physics, for which a wavepacket interpretation of (2) seems to facilitate a deeper understanding. I recall being puzzled by each of the following questions when I was an undergraduate physics major:

1. Why does $\boldsymbol{F} = md^2\boldsymbol{r}/dt^2$ involve the *second* derivative (and not, e.g., the *third*)?
2. Why do classical particles obey a variational principle?
3. Why do Bloch wave functions in a periodic lattice possess an *effective mass*?

I pondered the first question briefly as a freshman, and the other two questions arose in later years, as I was beginning to realize that such questions often lead nowhere. Nevertheless, it is always interesting when such questions do have answers.

My best answer to the third question is worth mentioning, because it involves both the converse to Goldstein's assertion, as well as some simple arguments that any physics student should be able to grasp. Since Block waves are described by a dispersion relation, one could use (2) to conclude that electrons often move as if they had a different "effective" mass. But is it true that wavepackets of **all** linear waves obey (2) in the eikonal limit? Wienberg's proved applies to wavepackets of coupled partial differential equations.[8] As the title of his article suggests, Weinberg did not claim to prove this for all plasma waves, because not all linear plasma waves are hydrodynamic[15,16]. Bloch wavefunctions[17] in a periodic lattice experience strong short-range potentials that also violate the assumptions used by Weinberg.

Fortunately, there are at least three simple cases where (2) is verified for any wave that exhibits linear superposition and interference.[18] Nothing else about the underlying wave equation is assumed. The simplest occurs in one-dimension, and when frequency is invariant: Take the differential, $d\omega = 0 = (\partial\omega/\partial x)dx+(\partial\omega/\partial k)dk$, and divide both sides by dt. Then use the easily proven fact that, $\partial\omega/\partial x$, is group velocity to obtain, $dk/dt=-\partial\omega/\partial x$.

To answer the second question, all particle motion governed by Hamiltonian equations (2), will also obey a variational principle.[10] Hence, both the *"traditional"* and the *"linearized"* approaches to General Relativity are equally strong in postulating either a Hamiltonian, or variational principle. The two postulates are equivalent, but incorporated differently in the two approaches. In the *traditional approach*, the concept of *path length* (and associated variational principle) is essential to Riemannian calculus. Since the *metric* is taken to be a fundamental field in this approach, the postulate that particles follow geodesics of this metric is natural. In the *linearized approach*, we must seek out a dispersion relation. In both approaches, the ultimate justification for the nature of particle motion is the existence of a quantum wave equation analogous to Schrödinger's.

The first, and most naïve of these three questions, concerns Newton's theory of motion, and can be answered using a non-relativistic argument. Reference 9 analyzes the spreading of wavepackets for dispersion relations of the form, $\omega=k^n$, and concludes that, n = 2, in free space. If the only possible non-relativistic free-particle dispersion relation is $\omega=Ak^2$, then the simplest possible wave equation for a non-zero potential is probably, $\omega=Ak^2+B\varphi$. Schrodinger's equation and its classical limit follow.

There is yet another reason to view particles as wavepackets in General Relativity. Although one could view $(-\omega, \mathbf{k})$ as *Hamiltonian-momentum*, Riemannian calculus naturally yields two vector fields that link $(-\omega, \mathbf{k})$ directly to *frequency-wavenumber*. One of these vectors describes location, or displacement, and is a *contravariant* vector, meaning that its components are written with superscripts. The other vector field involves the gradient, which acting upon a plane wave, will generate wavenumber: $\nabla e^{i\mathbf{k}\cdot\mathbf{r}} = i\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{r}}$. In four dimensions, the gradient generates *frequency-wavenumber*. This vector is *covariant* (subscripted).

A wavepacket is a function of the form, *ψ=A(**r**,t)exp(i**k·r**−iωt),* where the envelope, *A(**r**,t),* is a localized function with constant phase. The localization of the envelope, *A(**r**,t)* must be sufficiently small, so that we effectively treat the wavepacket as a particle. Yet, the wavepacket must be sufficiently large in extent that it contains many wavelengths. Such a wavepacket is sketched in Figure 2, with five dotted lines that represent crests in a region of space where they are of relatively large amplitude. Hence the wavepacket is about five wavelengths long, which by Heisenberg's uncertainty principle is marginially sufficient to localize momentum in k-space.

The wavepacket of Figure 2 is placed within an anisotropic lattice structure, as one might encounter in a crystal. The vector, $\boldsymbol{\xi} = \hat{\boldsymbol{e}}_{\boldsymbol{1}}\xi^1 + \hat{\boldsymbol{e}}_{\boldsymbol{2}}\xi^2$, represents a small displacement about some local origin. The lower right corner shows part of a "net" that can facilitate the labeling of ions with contravariant (superscripted) coordinates. The point (4, 2) shaded. The lattice structure exhibits a gentle inhomogeniety. There is virtually no loss of generality if the this inhomogeniety is small on a scale where the coordinates advance by one unit. This permits us to define a *coordinate basis vector* as approximately one "step", in which a variable advances by one unit, $\hat{\boldsymbol{e}}_{\boldsymbol{i}} \equiv \partial\boldsymbol{\xi}/\partial x^i \approx \Delta\boldsymbol{\xi}$, where Δx^i=1.

Figure 2 is a two dimensional analog that permits visualization of a simple, 2-dimensional version of Riemannian curvature, if one imagines that the net is placed over a smooth but uneven surface. The figure can also describe an actual coordinate system in space-time. Imagine that the net is three dimensional, and that the connecting strands, and circles, are essentially massless. There is no need for a simultaneous set of clocks at the circles, because the observer of this "universe" could use a single local clock to record the time-like coordinate for each event, which the observer presumably could see occur adjacent to one of the circles. (It should be noted that this paper does not label space-time in this fashion.)

It is instructive to appreciate the consequence of using non-orthogonal basis vectors when attempting to take inner products, but also note the elegance of a notation that automatically sums repeated indices and permits the metric to raise and lower indices:

$$\boldsymbol{\xi}\cdot\boldsymbol{\xi} = (\hat{\boldsymbol{e}}_{\boldsymbol{1}}\xi^1 + \hat{\boldsymbol{e}}_{\boldsymbol{2}}\xi^2)\cdot(\hat{\boldsymbol{e}}_{\boldsymbol{1}}\xi^1 + \hat{\boldsymbol{e}}_{\boldsymbol{2}}\xi^2) = \sum_{i,j=1}^{2}\xi^i(\boldsymbol{e_i}\cdot\boldsymbol{e_j})\xi^j = \xi_i g^{ij}\xi_j = \xi_i\xi^i$$

(3)

This can define the metric tensor as, $g_{ij} = \boldsymbol{e_i}\cdot\boldsymbol{e_i}$. The simplest non-trivial metric is associated with polar coordinates on a flat plane: $\Delta s^2=\Delta r^2+r^2\Delta\theta^2$. The associated 2×2 tensor is diagonal: $g_{ij}= diag(1,r^2)$.

The directions, but not magnitudes, of an entirely different pair of basis vectors are also shown in Figure 2. These are sometimes called the *dual basis*, and describe wavenumber: $\boldsymbol{k} = \underline{\boldsymbol{e}}^{\boldsymbol{1}}k_1 + \underline{\boldsymbol{e}}^{\boldsymbol{2}}k_2$, where, $\underline{\boldsymbol{e}}^{\boldsymbol{i}} \equiv \underline{\boldsymbol{\nabla}}x^i$, and the gradient is defined below.[19] Though the coordinate system is not orthogonal, the basis vectors are mutually orthogonal: $\underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\cdot\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}= 1$, if $\alpha = \beta$; otherwise it vanishes. This permits the inner product to take on the simple form associated with orthogonal coordinates: $\boldsymbol{k}\cdot\boldsymbol{\xi} = k_j\xi^j$. In the example of Fig. 2, the displacement, $\boldsymbol{\xi} = \hat{\boldsymbol{e}}_{\boldsymbol{1}}1 + \hat{\boldsymbol{e}}_{\boldsymbol{2}}3$, pierces exactly one wavelength. Therefore, $k_j\xi^j = -2\pi$, The magnitude is also intuitive: $k = (k_jk^j)^{1/2} = 2\pi/\lambda$, where the wavelength, λ, is the distance between two crests (dotted lines) in the figure.

In four dimensions, frequency-wavenumber is the (subscripted) covariant form, $K_\alpha = (-\omega, k_1 ,k_2, k_3)$. The perturbation of the metric will force curvilinear coordinates upon the theory. Nevertheless, we may always write a plane wave in the familiar form, $exp(i\boldsymbol{k}\cdot\Delta\boldsymbol{r}-i\omega\Delta t)= exp(iK_\alpha\xi^\alpha)$. But in a perturbed metric, the (superscripted) contravariant components, $K^\alpha\equiv g^{\alpha\beta}K_\beta$, will no longer represent frequency and wavenumber. Appendix B shows that K^α does have physical meaning, however.

The free particle Hamiltonian/dispersion relation can be expressed in one of three equivalent ways. Two are:

$$\omega^2 = m^2 + k^2$$

$$m^2 + K_\alpha\eta^{\alpha\beta}K_\beta = 0$$

(4ab)

The third utilizes diagonal 4×4 Minkowskian metric,

$$\eta^{\alpha\beta} = \eta_{\alpha\beta} = diag(-1,1,1,1),$$

(5)

to obtain, $m^2 + K_\alpha K^\alpha = 0$, an expression that will later be shown to hold, even in the nonlinear case of General Relativity. Writing the coordinate variables as,

$x^{\alpha}=(x^0, x^1, x^2, x^3) = (t, x, y, z)=(t, \mathbf{r})$, the proper time, Δτ, between two events accessible by a particle travelling at or less than the speed of light is,

$$-\mathrm{d}\tau^2 = \mathrm{d}x^{\alpha} g_{\alpha\beta} \mathrm{d}x^{\beta} = -\mathrm{d}t^{*2} + \mathrm{d}x^{*2} + \mathrm{d}y^{2*} + \mathrm{d}z^{*2} \leq 0\,, \tag{6}$$

where the asterisks (*) remind us that these are Cartesian variables that properly measure distance and time, but can only be defined locally.

III. Gradients in Riemannian Calculus

The postulated linearized field equation, $\partial_{\alpha}\partial^{\alpha}\Phi^{\mu\nu}=-16\pi T^{\mu\nu}$, must be reexamined when the perturbation of the metric, alluded to at (1), is achieved. Fortunately, this field equation remains unchanged in a first-order theory. The purpose of this section is to show the reader not yet proficient in Riemannian calculus why this is so.

Again it is helpful to illustrate principles in two dimensions. Using either a rotation matrix, or simply plotting simple points such as (1, 0) and (1, 1) it easily verified that the following transformation,

$$x = \frac{x' + y'}{\sqrt{2}} \qquad y = \frac{x' - y'}{\sqrt{2}} \tag{7}$$

is equivalent to a rotation of coordinates by 90 degrees. Partial derivatives in these coordinates are found using the chain rule. Let $\varphi= \varphi(x,y)$ be any scalar function, and use the chain rule to obtain, for example,

$$\frac{\partial\varphi}{\partial x'} = \frac{\partial\varphi}{\partial x}\frac{\partial \mathrm{x}}{\partial x'} + \frac{\partial\varphi}{\partial y}\frac{\partial \mathrm{y}}{\partial x'} = \frac{1}{\sqrt{2}}\left(\frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y}\right) \tag{8}$$

A second application of the chain rule yields the two identities,

$$\frac{\partial^2\varphi}{\partial x'\partial y'} = \frac{\partial^2\varphi}{\partial x^2} - \frac{\partial^2\varphi}{\partial y^2}$$

$$\frac{\partial^2\varphi}{\partial x'^2} + \frac{\partial^2\varphi}{\partial y'^2} = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} \tag{9ab}$$

In spite of "symmetric" appearances, the operator, $\partial^2/\partial x\partial y \equiv \partial_{xy}$, is not symmetric (or *invariant*) under rotation of coordinates. Since free space is assumed to have isotropic symmetry, the operator ∂_{xy} is not likely to appear in a fundamental two-dimensional field theory. On the other hand, $\partial_x^2 + \partial_y^2 \equiv \partial^i\partial_i$, does possess the required invariance. This is why the "del" operator,**∇**, is so useful.

In the Riemannian calculus of General Relativity, the gradient is:

$$\underline{\boldsymbol{\nabla}} \equiv \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\partial_\alpha = \sum_{\alpha=0}^{3} \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}} \ \partial/\partial x^\alpha \quad .$$

(10)

In an "empty universe", gravity and hence curvature are absent. Hence, we can use the 'flat' Minkowskian coordinate system, and this operator guarantees invariance under both rotation, as well as Lorentz transformations. In General Relativity, the invariance is even greater – all coordinate transforms are included, a fact that will be exploited in Sec. VII.

Let φ, $\mathbf{A} = \boldsymbol{e_\beta} A^\beta$, and $\overleftrightarrow{Q} = \boldsymbol{e_\beta}\, \boldsymbol{e_\gamma} Q^{\beta\gamma}$, be arbitrary scalar, vector, and tensor fields, respectively. Operate on each with the differential operator to obtain, $\underline{\boldsymbol{\nabla}}\varphi$, $\underline{\boldsymbol{\nabla}}\boldsymbol{A} = \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\partial_\alpha\left(\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\, A^\beta\right)$, and $\underline{\boldsymbol{\nabla}}\overleftrightarrow{\boldsymbol{Q}} = \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\partial_\alpha\left(\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\, \hat{\boldsymbol{e}}_{\boldsymbol{\gamma}}\, Q^{\beta\gamma}\right)$, respectively:

$$\underline{\boldsymbol{\nabla}}\varphi \equiv \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\partial_\alpha\varphi$$

$$\underline{\boldsymbol{\nabla}}\boldsymbol{A} = \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\, \partial_\alpha A^\beta \ + A^\beta\{\underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\partial_\alpha\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\}$$

$$\underline{\boldsymbol{\nabla}}\overleftrightarrow{\boldsymbol{Q}} = \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}}\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\, \hat{\boldsymbol{e}}_{\boldsymbol{\gamma}}\ \partial_\alpha Q^{\beta\gamma} \ + \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}} Q\left\{\left(\partial_\alpha\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\right)\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\right\} + \underline{\boldsymbol{e}}^{\boldsymbol{\alpha}} Q^{\beta\gamma}\left\{\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\, \partial_\alpha\hat{\boldsymbol{e}}_{\boldsymbol{\beta}}\right\}$$

(11abc)

A few comments are in order: First, all terms inside the curly brackets, { }, represent derivatives of basis vectors. Second, we see that differentiation increases the rank of a tensor. Third, it was necessary to change indices, e.g. from, $\mathbf{A}=\mathbf{e}_{\boldsymbol{\alpha}}\mathrm{A}^\alpha$, to $\mathbf{A}=\mathbf{e}_{\boldsymbol{\beta}}\mathrm{A}^\beta$, in order to avoid conflict with the summed index, α, of the differential operator, $\mathbf{e}^{\boldsymbol{\alpha}}\partial_\alpha$.

The terms in the curly brackets in (11bc) are called *connection coefficients*, and are not necessarily related to Riemannian curvature. They appear in expressions for **∇×A**, or **∇·A** in cylindrical and spherical coordinates, for example. One of the great triumphs of Riemannian calculus, is that Christoffel's *connection coefficients* can be calculated from the metric. One example will suffice, stated without proof:

$$\underline{\boldsymbol{\nabla}}\boldsymbol{A} = \left[\partial_\beta A^\alpha + \Gamma^\alpha_{\gamma\beta} A^\gamma\right]\hat{\boldsymbol{e}}_{\boldsymbol{\alpha}}\underline{\boldsymbol{e}}^{\boldsymbol{\beta}}. \tag{12}$$

From the box labeled *Riemannian Calculus* in the concept map, we see that the connection coefficients involve first derivatives of the metric. If the coordinate system is sufficiently close to a flat Minkowskian space, then derivatives of A^β, and $Q^{\beta\gamma}$ *are approximately* $\partial_\alpha A^\beta$, and $\partial_\alpha Q^{\beta\gamma}$, respectively, because connection coefficients appear as products of two first order terms (e.g., the product of A^μ and $\Gamma^\alpha_{\gamma\beta}$). Such terms are second order and can therefore be neglected in a linearized theory of gravity.

A useful rule of thumb is that if Riemannian calculus is handled skillfully, basis vectors are almost never written explicitly. This permits one to freely and quickly interchange the various expressions tensors such as $Q^{\alpha\beta}$, $Q_\alpha{}^\beta = g_\alpha{}^\delta Q^{\delta\beta}$, *or* $Q_{\alpha\beta} = g_{\beta\delta} Q_\alpha{}^\delta$. But skill at Riemannian calculus does not come easily. Consider for example, the differential of a vector, which has the merit of not raising the rank. Let $A^\alpha = A^\alpha(x^0, x^1, x^2, x^3)$ be a vector field, and consider two entirely different differentials that results from evaluating this field at two different places: One describes how the components of the vector change, $dA^\alpha = \xi^\beta \partial_\beta A^\alpha$, and is a simple application of the chain rule in the calculus of partial differentials. The other differential requires that vectors at two different locations be compared as if they had identical basis vectors: $DA^\alpha = dA^\alpha + \Gamma^\alpha_{\beta\gamma} A^\beta \xi^\gamma$. Letting, $A^\alpha = dx^\alpha/d\tau$ along the particle's path, this leads to the more direct derivation of geodesic motion alluded to in the second paragraph of Sec. I. These manipulations are difficult to learn, and even more difficult to fully grasp. Fortunately, much of the abstractness and complexity of Riemannian calculus is not required in this *linearized approach* to Relativity

IV. Constructing the "Linearized Approach"

We begin in flat Minkowskian space with the postulated field equation and Lorentz gauge condition:

$$\partial^\alpha \partial_\alpha \Phi^{\mu\nu} = -16\pi T^{\mu\nu}$$
$$\partial_\mu \Phi^{\mu\nu} = 0 \tag{13ab}$$

From electrodynamics[12], we know that a solution to (13a) is given by a Coulomb-like integral with $T^{\mu\nu}$ as the source: $\Phi^{\mu\nu} = 4\int\left[T^{\mu\nu}_{Ret}/r\right]d^3x$. The source is evaluated at the retarded[20] time, $T^{\mu\nu}_{Ret}(t) = T^{\mu\nu}(t-r)$.

The energy-momentum stress for a cold gas is $T^{\mu\nu} = \rho U^{\mu}U^{\nu}$, where ρ is mass density in the fluid's reference frame and $U^{\mu} = dX^{\mu}/d\tau$ is the 4-velocity.[21] A gas is *cold* if the atoms are in motion, but undergo no relative motion as they move. Since, $U = (1, 0, 0, 0)$, for a particle at rest, $\rho U^{\mu}U^{\nu}$ is the most obvious way to translate mass density into a rank-2 tensor. The static case of motionless gravitating mass is calculated as follows: Since $T^{00} = \rho$ is the only non-zero element, and the solution to (13a) is,

$$\Phi^{\alpha\beta} = \mathrm{diag}(-4V, 0,0,0), \tag{14}$$

where $\nabla^2 V = 4\pi\rho$, and $V << 1$ in the weak field limit. The Minkowskian metric (5) permits us to raise and lower indices, provided we ignore higher order terms. Hence to first order, $\Phi^{00} = \Phi_{00} = -4V$. The trace (or contraction) is written with repeated indices, so that $\Phi = \Phi^{\alpha}{}_{\alpha} = +4V$. Note that the trace of the Minkowskian metric is, $\eta = 4$

Next we find a dispersion relation, taking space-time to be Minkowskian, as we postpone the issue of metric and the gravitational redshift of light. (When the metric is perturbed, it will emerge that this dispersion relation remains intact.) Our search for an appropriate dispersion relation is facilitated by the fact that a particle's path is independent of its mass. This restricts the dispersion relation to an expression of the form:

$$0 = m^2 + K_\alpha \eta^{\alpha\beta} K_\beta + b_1 \Phi K_\alpha \eta^{\alpha\beta} K_\beta - b_2 K_\alpha \Phi^{\alpha\beta} K_\beta\ . \tag{15}$$

The first two terms are the free-particle dispersion relation, and constants (b_1, b_2) on the two first-order correction terms must be found to complete the theory. To understand why only two terms need be included in (15), consider the free-particle relation, $\omega^2 = m^2 + k^2$. If the mass is doubled, while speed is kept constant, then each term increases by a factor of four. Hence, all terms in the free-particle dispersion relation are quadratic in the variables (m, k, ω), and we expect the same for terms that are added as we to account for gravity. Otherwise the influence of gravity on motion will become dependent on mass.

Of these three quantities (*m, k, ω*), two of them (*ω*, k) will always appear together as frequency-wavenumber: $K_\alpha \equiv (-\omega, \mathbf{k})$. Furthermore, we are at liberty to refrain from using the parameter, *m,* because to first order in field strength, the approximation, $m^2 \approx -K_\alpha \eta^{\alpha\beta} K_\beta$, is sufficient.

To contract a tensor is to take an inner product, or to sum two identical indices, one covariant (subscripted) and the other contravariant (superscripted). The result is a tensor reduced in rank by two. Hence, all terms in (15) contain two contractions of the rank-4 tensor, $\mathbf{K}\,\mathbf{K}\overleftrightarrow{\Phi}$. The symmetry of, $\Phi^{\alpha\beta} = \Phi^{\beta\alpha}$, stipulates that there are only two ways to contract something like $\Phi_{\square\square} K_{\square} K_{\square}$ down to a rank-2 tensor. These two possible contractions are shown in (15).

V. Finding three "easy" constraints and perturbing the metric

Two more constants will appear as part of the linearized theory, and with four unknown constants, we shall require four physical constraints. Each constraint will yield a single equation, but only three of them are easily understood. They are:

1. Newtonian motion at low speeds in a static gravitational field.
2. The gravitational redshift associated with the wave of a massless particle.
3. The invariant speed of light (which equals unity in these units).

A theoretical argument for the fourth constant is found in Appendix B. But, the reader may choose to instead rely on experimental evidence concerning the deflection of light by the Sun's gravity.

First we recover Newtonian motion by rewriting (15) in the static case. Using $\Phi = 4V$, $K_\alpha \eta^{\alpha\beta} K_\beta = -m^2$, and $\Phi^{00} = -4V$, to obtain $K_\alpha \Phi^{\alpha\beta} K_\beta = -4V\omega^2$, we have:

$$\begin{aligned} \omega^2 &= m^2 + k^2 - 4Vb_1 m^2 + 4Vb_2\omega^2 \\ &= m^2 + k^2 + 4(b_2 - b_1)m^2 V + 4b_2 k^2 V \end{aligned} \tag{16ab}$$

The last step used the zeroth order approximation, $\omega^2 \approx m^2 + k^2$, on a term that is already small (first order). Newtonian motion is obtained by writing the Schrödinger's dispersion relation, but with an extra constant term to include a particle's rest energy. Fundamental constants equal to unity are temporarily inserted for clarification:

$$\hbar\omega = mc^2 + \frac{1}{2m}\hbar^2 k^2 + mV$$
$$\omega = m + \frac{1}{2m}k^2 + mV.$$

(17ab)

The first term on the RHS of (17) is much larger than the other two terms in the non-relativistic limit ($v^2 \approx k/m \ll 1$). In a linearized theory, we also drop terms quadratic in the field strength, V. The result is that when both sides of (17) are squared, the result, in the Newtonian limit, is:

$$\omega^2 = m^2 + k^2 + 2m^2 V + small\ terms$$

(18)

Neglect$4b_2k^2V$ in (16b) at low speeds, and compare with (18). This establishes the first constraint on our four unknown constants:

$$2b_2 - 2b_1 = 1.$$

(19)

Next we find the dispersion relation for light by letting rest-mass vanish in (16) to obtain, $\omega^2 = k^2(1 + 4Vb_2)$. This can be simplified using, V<<1, and the expansion, $(1+\varepsilon)^n \approx 1 + n\varepsilon$, for small ε:

$$\frac{\omega}{k} = 1 + 2b_2V.$$

(20)

This does not imply that $b_2 = 0$, because the gravitational redshift introduces a perturbation of the metric. Hence, (20) does not represent the perceived speed of light. In the static case, we perturb the metric by introducing two additional constants, b_3, and b_4:

$$dt^* = (1 + b_3V)dt$$
$$ds^* = (1 + b_4V)ds,$$

(21ab)

where $ds^2 \equiv dx^2 + dy^2 + dz^2$. Following the discussion at (1) we set , $\omega^*\,\Delta t^* = \omega^*\,(1 + b_3V)\Delta t = \omega\,\Delta t$. Hence, to first order in field strength, (21) is equivalent to:

$$\omega^* = (1 - b_3V)\omega$$
$$k^* = (1 - b_4V)k.$$

(22ab)

As discussed in Sec. I, the gravitational redshift is, $\Delta\omega = -\omega V = \omega^* - \omega$. From (22a) we have our second equation involving$(b_1, b_2, b_3, b_4\ )$:

$$b_3 = 1. \tag{23}$$

Our third physical constraint arises from the perceived speed of light, which must equal unity in our units:

$$\frac{\omega^*}{k^*} = \frac{(1 - b_3 V)}{(1 - b_4 V)}(1 + 2b_2 V) \approx 1 - b_3 V + b_4 V + 2b_2 V = 1. \tag{24}$$

Hence,

$$b_3 = 2b_2 + b_4. \tag{25}$$

The differentials (21ab) represent the metric perturbation in an unconventional fashion. So, we square and drop terms in (21ab) that are quadratic in V to obtain:

$$-d\tau^2 = -dt^{*2} + ds^{*2} = (-1 - 2b_3 V)dt^2 + (1 + 2b_4 V)ds^2. \tag{26}$$

To complete this perturbation of the metric, we must generalize from the static case to the dynamic case of gravitating masses in motion. In a linearized (weak field) theory, it is conventional to take the new metric as almost Minkowskian, defining the metric perturbation, $h_{\alpha\beta}$, as follows:

$$\begin{aligned} g_{\alpha\beta} &= \eta^B_{\alpha\beta} + h_{\alpha\beta} \\ g^{\alpha\beta} &= \eta_T^{\alpha\beta} - h^{\alpha\beta}. \end{aligned} \tag{27}$$

Now, $g_{\alpha\beta}$ and its inverse $g^{\alpha\beta}$, define the true metric. In a first order theory, $h_{\alpha\beta}$ <<1. The *two* Minkowskian metrics have constant coefficients,

$$\eta^B_{\alpha\beta} = diag(-1,1,1,1) = \eta_T^{\alpha\beta}, \tag{28}$$

but are neither truly constant tensors nor precisely equal to each to each other. The approximation $\overleftrightarrow{\eta}_T \approx \overleftrightarrow{\eta}_B$ is sufficient everywhere in this paper, and the distinction is mentioned solely to keep (27) from seeming nonsensical: Riemannian calculus stipulates that, $A_{\alpha\beta}+B_{\alpha\beta}=C_{\alpha\beta}$ if and only if $A^{\alpha\beta}+B^{\alpha\beta}=C^{\alpha\beta}$.

Equation (26) is consistent with a diagonal metric that can be partitioned and multiplied as follows:

$$-\mathrm{d}\tau^2 = \mathrm{d}x^\alpha\big(\eta_{\alpha\beta} + h_{\alpha\beta}\big)\mathrm{d}x^\beta = [dt \quad dx^j]\begin{bmatrix}-1 + h_{00} & 0 \\ 0 & 1 + h_{jj}\end{bmatrix}\begin{bmatrix}dt \\ dx^j\end{bmatrix}$$
$$= -(1 - h_{00})dt^2 + (1 + h_{11})ds^2\,. \tag{29}$$

This is only valid in the static case. But comparison with (26) establishes that $\mathrm{h}_{00} = -2\mathrm{b}_3\mathrm{V}$ and and $\mathrm{h}_{11} = \mathrm{h}_{22} = \mathrm{h}_{33} = 2\mathrm{b}_4\mathrm{V}$. Since the static-case metric perturbation is linear in static potential, V, we postulate that the dynamic metric perturbation is a linear combination of $\Phi_{\alpha\beta}$ and $\Phi\eta_{\alpha\beta}$. Using Φ=4V, $\Phi^{\alpha\beta}$= diag(−4V, 0, 0, 0), and $\eta^{\alpha\beta}$= diag(−1, 1, 1, 1), it is easily verified that,

$$\mathrm{h}_{\alpha\beta} = \frac{1}{2}(\mathrm{b}_3 - \mathrm{b}_4)\Phi_{\alpha\beta} + \frac{1}{2}\mathrm{b}_4\Phi\eta_{\alpha\beta}\,, \tag{30}$$

is the only linear combination of $\overleftrightarrow{\Phi}$ and $\overleftrightarrow{\eta}\Phi$ that recovers (26) in the static case. We therefore postulate (30) to hold in the dynamic case, but must also recognize the following consequence: Equation (30) can be solved to express $\overleftrightarrow{\Phi}$ as a function of $\overleftrightarrow{\mathrm{h}}$. This permits us to consider the metric as the sole field that governs particle motion, and we are at liberty to "demote" $\overleftrightarrow{\Phi}$ to nothing more than a convenient expression involving the metric. The *traditional* and *linearized* approaches are beginning to merge!

The fourth constraint is so problematical that we pause to consolidate our theory so that it contains a single unknown parameter, *b*, defined so that the 'correct' value is *b=1*. Equations (19), (23), and (25) combine to yield:

$$b_1 = \frac{b}{2}, \quad b_2 = \frac{1 + b}{2}, \quad b_3 = 1, \quad b_4 = -b. \tag{31}$$

Substitute into (30), and take the trace. This yields, $2h = (1 - 3b)\Phi$, and permits us to express the field in terms of the metric. Hence:

$$h_{\alpha\beta} = \left(\frac{1+b}{2}\right)\Phi_{\alpha\beta} - \frac{b}{2}\Phi\eta_{\alpha\beta}$$

$$\Phi_{\alpha\beta} = \left(\frac{2}{1+b}\right)h_{\alpha\beta} + \frac{2bh}{(1+b)(1-3b)}\eta_{\alpha\beta}.$$

(32ab)

The linearized dispersion relation can be expressed either in terms of $\overleftrightarrow{\Phi}$ or $\overleftrightarrow{\mathrm{h}}$:

$$0 = m^2 + K_\alpha\eta^{\alpha\beta}K_\beta + \frac{b}{2}\Phi K_\alpha\eta^{\alpha\beta}K_\beta - \frac{1+b}{2}K_\alpha\Phi^{\alpha\beta}K_\beta$$

$$0 = m^2 + K_\alpha\left(\eta^{\alpha\beta} - h^{\alpha\beta}\right)K_\beta = m^2 + K_\alpha K^\beta.$$

(33ab)

Equation (33b) exhibits remarkable simplicity for a dispersion relation describing wavepacket motion under the influence of gravity. As one might guess, it represents geodesic motion, as is verified in Appendix A.

While Appendix B shows that, $b = 1$, is necessary to make the theory invariant under coordinate transformations, it is more illuminating to establish this from experimental evidence involving the deflection of light by the Sun. The motion of light and particles, in the presence of a static gravitational field, is discussed in the next section.

VI. The breakdown of Euclidean Geometry

Here we derive equations of motion for the special case of a static field. No assumption about the value of b is made, and we shall see that the breakdown of Euclidean geometry occurs whenever, $b \neq 1$. We shall also see that b is closely related to the deflection of light directed perpendicular to gravity. Hence, this section permits one to use experimental evidence concerning the deflection of light by the Sun to establish, $b \approx 1$ (within experimental error).

From and (16b) and (31), the dispersion relation for a static field is

$$\omega = \sqrt{m^2 + k^2 + 2m^2V + (2+2b)k^2V}\,,$$

(34)

Where, $m=0$, for light. Take the gradient with respect to $\mathbf{k}$ to obtain the *group velocity*:

$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} = \frac{\partial\omega}{\partial\boldsymbol{k}} = \frac{\boldsymbol{k}}{\omega}(1 + 2V + 2bV) \approx \frac{\boldsymbol{k}}{\omega} .$$

(35)

The zeroth order approximation, $\boldsymbol{v} \approx \boldsymbol{k}/\omega$, may be used, but with care. For example, upon taking the time derivative of (35) to obtain acceleration, we keep all terms that are first order in V. Using $\mathrm{dV/dt} = \boldsymbol{v} \cdot \boldsymbol{\nabla} V$, and taking the time derivative of (35) we obtain,

$$\boldsymbol{a} = \frac{d^2\boldsymbol{r}}{dt^2} \approx \frac{1}{\omega}\frac{d\boldsymbol{k}}{dt} + \frac{\boldsymbol{k}}{\omega}(2 + 2b)\boldsymbol{v} \cdot \boldsymbol{\nabla} V .$$

(36)

Note that, $d\boldsymbol{k}/dt$ is also first order in field strength,

$$\frac{d\boldsymbol{k}}{dt} = -\frac{\partial\omega}{\partial\boldsymbol{r}} = -\frac{m^2 + k^2 + bk^2}{\omega}\boldsymbol{\nabla} V .$$

(37)

This can be simplified using the zeroth order approximation, $\boldsymbol{k} \approx \omega\boldsymbol{v} \approx \gamma m\boldsymbol{v}$, where $\gamma = (1 - v^2)^{1/2}$. The result is,

$$a = -(1 + bv^2)\boldsymbol{\nabla} V + (2 + 2b)\boldsymbol{v}\boldsymbol{v} \cdot \boldsymbol{\nabla} V ,$$

(38)

where *mass* does not appear, as one would expect for a theory where a particle's path does not depend on this parameter. For light, we set $v = 1$.

The importance of metric and coordinate system in the interpretation of (38) cannot be overstated. Consider, for example, a simple but non-physical example that is analogous to the parallel plate capacitor in electromagnetism.[22] Without negative mass, we can have only one plate. To maintain V<<1, consider a disk, and avoid fringing fields by assuming that we are very close to the center of the disk. The disk is shown at the top of Figure 3. The two characters posing as sentient beings on each side of the disk are not drawn to scale. By analogy with electrostatics, we have for y>0:

$$V = a_g y ,$$

(39)

where,

$$a_g = 2\pi G \frac{Mass}{Area} \approx 9.8 \frac{m}{s^2}, \tag{40}$$

if the surface mass density of the "planet" is $2.34 \times 10^{10}\ kg/m^2$. The diameter of this disk must be much less than one light-year, since $1/a_g = c^2/a_g \approx 10^{16}\ m$, in order to ensure that, V<<1, everywhere.

Even though the disk appears 'flat' in x-y coordinates, each observer perceives it to be concave, if *b>1*. The unit cubes that fill xyz space will not be perceived as cubes, since (21b) stipulates that the top lengths are shorter. Such non-cubical blocks cannot fit snugly into a Euclidean universe. If *b=1*, the radius of curvature is $1/a_g$, or about one light year for Earth-like gravity. Note that the linear theory breaks down just as the disk begins to curve into a sphere.

If (39) is substituted into (38), it is found that light directed horizontally in the x-direction obeys, $d^2y/dt^2 = -(1+b)a_g$. The so-called "anomalous" extra downward acceleration, $-ba_g$, represents the fact that the curved planet rises to meet the light. Consider a very simple analogy to horizontally directed light on earth. Neglecting atmospheric refraction and the immeasurable gravitational deflection by Earth's gravity, a horizontally directed beam of light would travel in a straight line. But it would "rise" in altitude due to the fact that the Earth is spherical. But this "rise" is entirely an artifact of a spherical coordinate system and the use of altitude to denote position.

Similarly, vertically directed light obeys, $d^2y/dt^2 = (2+2b)a_g$. This so-called vertical "acceleration" simply ensures that the perceived speed of light remains constant: $dy^*/dt^* = 1$. Hence both the "anomalous" deflection of horizontal light, as well as the upward "acceleration" of vertically directed light are artifacts of the coordinate system. The perceived motion of light is exactly what intuition would predict: There is no change in speed if the light is directed parallel to gravity, and the light seems to accelerate downward as do all falling bodies if the light is directed horizontally. A free-falling local observer will always observe light to be travelling in a straight line.

Euclidean geometry is valid if and only if *b=1*. One can rigorously establish this by comparing the radius and circumference of a sphere of uniform mass density. But Figure 3 demonstrates the breakdown more directly. If technological civilizations evolved independently on both sides of this "planet", they might observe the deflection of light and measure the bowel shaped curvature of the planet without discovering General Relativity. But the inhabitants of this planet would be surprised when they finally drilled holes though the "planet", measured its thickness, and concluded that part of the universe is "missing". Incidentally, this also illustrates the dangers of relying on intuitive or "physical" arguments in General Relativity. Without *a priori* knowledge of Einstein's relativity, one might invoke Euclid to construct a convincing (but false) physical argument that b must vanish.

Since this is a first order theory, we expect the equations of motion to represent first order corrections to uniform motion in an empty Minkowskian universe. This is evident for the light, since solutions to (38) deviate only slightly from, $\boldsymbol{r} = \boldsymbol{r_0} + \boldsymbol{v}t$, whenever, $v = |\boldsymbol{v}| = 1$. But consider solutions to (38) the low-speed Newtonian limit, $\boldsymbol{a} = -\boldsymbol{\nabla} V$. How can the Earth's path around the Sun be nearly geodesic? The answer is that the path through space-time of a non-relativistic (slow) particle is almost exclusively through the time-like coordinate, $-d\tau^2 \approx -dt^2$. The circular motion of a non-relativistic particle is merely a gentle helical "twist" in two spatial dimensions, of what is essentially a "straight-line" through t.

VII. Curvature at Last!

Let us informally define "curvature" as something that vanishes in an empty (flat) universe, and stipulate that it must vanish under any transformation (change of variables) within this space. We may transform to Cartesian-like variables, $x^{\alpha *} = (t^*, x^*, y^*, z^*)$, because they do exist in an empty universe. *Curvature* must therefore vanish for all coordinate transformations:

$$x^{\alpha *} = f^{\alpha}(x^0, x^1, x^2, x^3) \equiv f^{\alpha}(x^{\beta}) \tag{41}$$

Restricting ourselves to infinitesimal coordinate transformations, we consider transformations of the form:

$$x^{\alpha *} = x^{\alpha} + \xi^{\alpha} \tag{42}$$

Here, $\xi^{\alpha} = \xi^{\alpha}(x^0, x^1, x^2, x^3)$ is a differential vector field that is sufficiently small that the coordinates x^{α} and $x^{\alpha *}$are nearly identical.[23,24] Appendix B shows that the metric perturbation associated with this transformation is,

$$h_{\alpha\beta} = \partial_\alpha \xi_\beta + \partial_\beta \xi_\alpha \tag{43}$$

Equation (13a) suggests that the LHS of the field equation, $\partial^\alpha \partial_\alpha \Phi^{\mu\nu}$, involves second derivatives of the metric. Hence we seek a rank-4 tensor of the form:

$$Q_{\alpha\beta\gamma\delta} = \sum_{n=1}^{6} C_n \, \partial_{\square\,\square} h_{\square\,\square} \, . \tag{44}$$

The symbol, $\square$, represents a place mark for one of the four indices, $(\alpha, \beta, \gamma, \delta)$, where each index is used one and only one time. There are six terms in (44) because the combination of 4 objects taken 2 at a time is 4!/(2!2!)=6. If curvature always vanishes in an empty universe, then $Q_{\alpha\beta\gamma\delta} = 0$. To find the six coefficients, C_n, we solve:

$$0 = C_1 \partial_{\square\square} h_{\alpha\beta} + C_2 \partial_{\square\square} h_{\alpha\gamma} + C_3 \partial_{\square\square} h_{\alpha\delta} + C_4 \partial_{\square\square} h_{\beta\gamma} + C_5 \partial_{\square\square} h_{\beta\delta} + C_6 \partial_{\square\square} h_{\delta\gamma} \tag{45}$$

It is not necessary to fill in the blanks on the $\partial_{\square\,\square}$ operator because those subscripts are uniquely deduced from the subscripts on $h_{\square\,\square}$. When (43) is substituted into (45), we obtain,

$$0 = C_1 \partial_{\square\square\square} \xi_\alpha + C_1 \partial_{\square\square\square} \xi_\beta + C_2 \partial_{\square\square\square} \xi_\alpha + C_2 \partial_{\square\square\square} \xi_\gamma + C_3 \partial_{\square\square\square} \xi_\alpha + C_3 \partial_{\square\square\square} \xi_\delta + C_4 \partial_{\square\square\square} \xi_\beta + C_4 \partial_{\square\square\square} \xi_\gamma + C_5 \partial_{\square\square\square} \xi_\beta + C_5 \partial_{\square\square\square} \xi_\delta + C_6 \partial_{\square\square\square} \xi_\delta + C_6 \partial_{\square\square\square} \xi_\gamma \, . \tag{46}$$

Setting to zero the sum of all terms proportional to ξ_α, implies that $C_1 + C_2 + C_3 = 0$. Three other equations are obtained in the same way: $0 = C_1 + C_4 + C_5 = C_1 + C_4 + C_5 = C_2 + C_4 + C_6 = C_3 + C_5 + C_6$. With 4 independent homogeneous linear equations in 6 unkowns, we have a 2 dimensional subspace of independent solutions. After a bit of algebra all but C_1 and C_2 can be eliminated: $C_3 = -C_1 - C_2;\ C_4 = -C_1 - C_2;\ C_5 = C_2;\ C_6 = C_1$. Writing $Q_{\alpha\beta\gamma\delta}$ as a linear combination of two tensors we have an expression that must vanish in an empty universe:

$$Q_{\alpha\beta\gamma\delta} = C_1 Q^A_{\alpha\beta\gamma\delta} + C_2 Q^B_{\alpha\beta\gamma\delta} \tag{47}$$

where

$$Q^A_{\alpha\beta\gamma\delta} = \partial_{\square\square} h_{\alpha\beta} - \partial_{\square\square} h_{\alpha\delta} - \partial_{\square\square} h_{\beta\gamma} + \partial_{\square\square} h_{\delta\gamma}$$

$$Q^B_{\alpha\beta\gamma\delta} = \partial_{\square\square} h_{\alpha\gamma} - \partial_{\square\square} h_{\alpha\delta} - \partial_{\square\square} h_{\beta\gamma} + \partial_{\square\square} h_{\beta\delta}$$

(48)

Interchanging subscripts shows that both tensors are transposes of the linearized Riemann's curvature tensor: $Q^{A}_{\alpha\delta\beta\gamma} = 2R^{linear}_{\alpha\beta\gamma\delta}$, and $Q^{B}_{\alpha\beta\delta\gamma} = 2R^{linear}_{\alpha\beta\gamma\delta}$, where $R^{linear}_{\alpha\beta\gamma\delta} = \partial_\gamma \Gamma_{\alpha\beta\delta} - \partial_\delta \Gamma_{\alpha\beta\gamma}$, as can be verified from the concept map.

Though weak and primitive, this introduction to curvature does establish uniqueness: In an empty universe, the only linear rank-4 tensor, that must vanish and involves second derivatives of the metric, is either *linearized Riemann*, or a transpose thereof.

Appendix A: Geodesics and Debroglie's relations from the dispersion relation.

The purpose of this section is to establish that two Lagrangians are equivalent. One arises from classical Hamiltonian dynamics[10], which we shall write as, $L = \boldsymbol{k} \cdot \boldsymbol{v} - \omega$, where the dispersion relation contains the metric: $0 = m^2 + K_\alpha K^\alpha$. The other Lagrangian arises from the geodesic postulate[1] and variational principle, $\delta \int d\tau = 0$. To express this variation as an integral over the timelike variable, $t \equiv x^0$, write it as, $\int d\tau = \int \mathcal{L} dt$.

The dispersion relation can be expressed as, $F = 0 = m^2 + K_\alpha g^{\alpha\beta} K_\beta$, where F is a function of eight variables: $F = F(\omega, \boldsymbol{k}, t, \boldsymbol{r})$. Partial derivatives among these variables are obtained implicitly[25]: $dF = 0 = (\partial F/\partial z^1)dz^1 + (\partial F/\partial z^2)dz^2 + \dots + (\partial F/\partial z^8)dz^8$, where z^j represent the eight variables. Hence, for $i \neq j$, we have $\partial z^i/\partial z^j = -(\partial F/\partial z^i)/(\partial F/\partial z^j)$. For example

$$\frac{\partial \omega}{\partial k_j} = \frac{-\partial F/\partial k_j}{\partial F/\partial \omega} = \frac{-2g^{j\alpha}K_\alpha}{-2g^{0\alpha}K_\alpha} = \frac{K^j}{K^0}.$$

(A1)

From classical dynamics[10], the Lagrangian obeys the variational principle, $\delta \int L dt = 0$. It is related to the Hamiltonian (frequency) by:

$$L = k_j \frac{\partial \omega}{\partial k_j} - \omega = \frac{K_j K^j}{K^0} + K_0 = -\frac{m^2}{K^0},$$

(A2)

Next consider the geodesic postulate associated with the *traditional approach* to General Relativity. Let $\int d\tau = \int (d\tau/dt)dt$, so that,

$$\mathcal{L} = \frac{d\tau}{dt} = \sqrt{-\frac{dx^\alpha}{dt} g_{\alpha\beta} \frac{dx^\beta}{dt}}\,. \tag{A3}$$

The 16 terms in the sum can be grouped into 1 + 6+9 terms:

$$-\mathcal{L}^2 = \frac{dx^0}{dt} g_{00} \frac{dx^0}{dt} + 2\frac{dx^0}{dt} g_{0i} \frac{dx^i}{dt} + \frac{dx^i}{dt} g_{ij} \frac{dx^j}{dt} \tag{A4}$$

Using $dx^j/dt = \partial\omega/\partial k_j$, and $dx^0/dt = 1$, we have:

$$-\mathcal{L}^2 = g_{00} + 2g_{0i}\frac{K^j}{K^0} + \frac{K^i}{K^0} g_{ij} \frac{K^j}{K^0} = \frac{K^\alpha g_{\alpha\beta} K^\beta}{(K^0)^2} = -\left(\frac{m}{K^0}\right)^2 \tag{A5}$$

Hence, $L = m\mathcal{L}$, and the two Lagrangians are essentially the same.

Finally we obtain the physical meaning of the contravariant (superscripted) form of frequency-wavenumber. From (2) and (A1) we have, $dx^\alpha/dt = K^\alpha/K^0$. Hence,

$$\left(\frac{d\tau}{dt}\right)^2 = -\frac{dx^\alpha}{dt} g_{\alpha\beta} \frac{dx^\beta}{dt} = -\frac{K^\alpha}{K^0} g_{\alpha\beta} \frac{K^\beta}{K^0} = \frac{-K_\alpha K^\alpha}{(K^0)^2} \tag{A6}$$

Hence, $d\tau/dt = m/K^0$. Inserting Plank's constant, this can be written to formally resemble both of De Broglie's relations:

$$m\frac{dx^\alpha}{d\tau} = \hbar K^\alpha\,. \tag{A7}$$

Appendix B: Plausibility argument that b=1

This "proof" that, *b=1*, is based on a simple and obvious fact: a change of variables cannot create energy-momentum out of empty space. The argument is marred by a mathematical difficulty associated with a potential that diverges at far from the origin and therefore violates the assumption of weak fields. The discussion that ensues is not intended for the casual reader, and is included to establish that one could, in principle, recover linearized General Relativity without appealing to experimental data concerning the deflection of light by gravity.

First we establish (43), which defines the metric for any infinitesimal coordinate transformation. From (42) we take the differential and use the chain rule to first order, $d\xi^{\alpha} = dx^{\mu}\partial_{\mu}\xi^{\alpha}$

$$dx^{\alpha *} = dx^{\alpha} + dx^{\mu}\partial_{\mu}\xi^{\alpha}$$

(B1)

The metric perturbation is defined so that:

$$-d\tau^2 = dx^{\alpha}(\eta_{\alpha\beta} + h_{\alpha\beta})dx^{\beta} = (dx^{\alpha} + dx^{\mu}\partial_{\mu}\xi^{\alpha})\eta_{\alpha\beta}(dx^{\beta} + dx^{\nu}\partial_{\nu}\xi^{\beta})$$

(B2)

Hence, to first order,

$$\begin{aligned} dx^{\alpha}h_{\alpha\beta}dx^{\beta} &= dx^{\alpha}\eta_{\alpha\beta}\partial_{\nu}\xi^{\beta}dx^{\nu} + dx^{\mu}\partial_{\mu}\xi^{\alpha}\eta_{\alpha\beta}dx^{\beta} \\ &= dx^{\alpha}\partial_{\nu}\xi_{\alpha}\, dx^{\nu} + dx^{\mu}\partial_{\mu}\xi_{\beta}dx^{\beta} \end{aligned}$$

(B3)

Equation (43) results when the summed indices are interchanged:$\beta \leftrightarrows \nu$ on the first term on the LHS, and $\alpha \leftrightarrows \mu$ on the second term.

Consider an empty universe in which $\overleftrightarrow{T}$ vanishes everywhere. Use (31b) to express the field equations (13ab) in terms of the perturbed metric:

$$\frac{2}{b+1}\partial^{\alpha}\partial_{\alpha}\left(h_{\mu\nu} + \frac{bh}{1-3b}\eta_{\mu\nu}\right) = 0$$

$$\frac{2}{b+1}\partial^{\mu}\left(h_{\mu\nu} + \frac{bh}{1-3b}\eta_{\mu\nu}\right) = 0$$

(B4ab)

Use (43) to generate the identities, $\partial^{\mu}h_{\mu\nu} = \partial^{\mu}\partial_{\mu}\xi_{\nu} + \partial^{\mu}\partial_{\nu}\xi_{\mu}$, and $h = 2\partial_{\alpha}\xi^{\alpha}$, so that the gauge condition (B4b) implies:

$$\partial^2\xi_{\mu} = \partial_{\alpha}\partial^{\alpha}\xi_{\mu} = \frac{1-b}{3b-1}\partial_{\mu}\partial^{\beta}\xi_{\beta} = \frac{1-b}{3b-1}\partial_{\mu}(\partial \cdot \xi)$$

(B5)

To avoid excessive indices, it is helpful to write $\partial^{\alpha}\partial_{\alpha}$ as ∂^2, and $\partial^{\beta}\xi_{\beta}$as $(\partial \cdot \xi)$. Substitute this into(B4a), and the field equation becomes:

$$\frac{2}{b+1}\partial^{\alpha}\partial_{\alpha}\left(\partial_{\mu}\xi_{\nu} + \partial_{\nu}\xi_{\mu} + \frac{2b\eta_{\mu\nu}}{1-3b}\partial^{\beta}\xi_{\beta}\right) = 0$$

(B6)

Note that, $\partial^\alpha \partial_\alpha \equiv \partial^2$, commutes with the other operators, so that (B6) can be written as,

$$\frac{2}{b+1}\left(\partial_\mu \partial^2 \xi_\nu + \partial_\nu \partial^2 \xi_\mu + \frac{2b\eta_{\mu\nu}}{1-3b}\partial^\beta \partial^2 \xi_\beta \right) = 0\,.$$

(B7)

Use (B5) and the notation that follows, to replace ∂^2 by a term proportional to $(\partial \cdot \xi)$. The field equation can now be written as:

$$\left(\frac{4}{b+1}\right)\left(\frac{1-b}{3b-1}\right)\left\{\partial_\mu \partial_\nu + \frac{b\eta_{\mu\nu}}{1-3b}\partial^2\right\}(\partial \cdot \xi) = 0$$

(B8)

Note the factor of $(1-b)$. We can stipulate that, b=1, if we can find a transformation with two properties. First the operator in curly brackets { } must not cause the LHS of (B8) to vanish. Second, the transformation must satisfy the Lorentz gauge condition (B5). One such function is:

$$\xi^\alpha = \left[xyt, \quad \frac{1}{2}\frac{b-1}{3b-1}t^2 y, \quad \frac{1}{2}\frac{b-1}{3b-1}t^2 x, \quad 0\right].$$

(B9)

The metric generated by (B9) is not small everywhere. Nevertheless, the argument holds in a small neighborhood, so any transformation that approaches (B9) near the origin should achieve the same result.

To verify that (B9) forces, b=1, in (B8) we operate as follows:

$$(\partial \cdot \xi) = \partial_t(xyt) = xy$$

$$\partial^2 \xi_\beta = \left(-\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_x^2\right)\xi_\beta = \left[0, \quad \frac{1-b}{3b-1}y, \quad \frac{1-b}{3b-1}x, \quad 0\right]$$

$$\partial_\mu(\partial \cdot \xi) = [0, \quad y, \quad x, \quad 0]$$

(B10abc)

Equation (B10a) establishes that $\partial^2(\partial \cdot \xi) = 0$. But since $\partial_x \partial_y(\partial \cdot \xi) = 1$, we see that the only factor in (B8) that can force the LHS to vanish is (b−1). And substitution into (B5) establishes that the Lorentz gauge condition would be satisfied, even if b did not equal unity. Therefore, (B9) represents a coordinate transformation that satisfies the Lorentz gauge condition, for which the field equation (B8) would stipulate that energy-momentum stress is created in an empty universe…unless *b=1*.

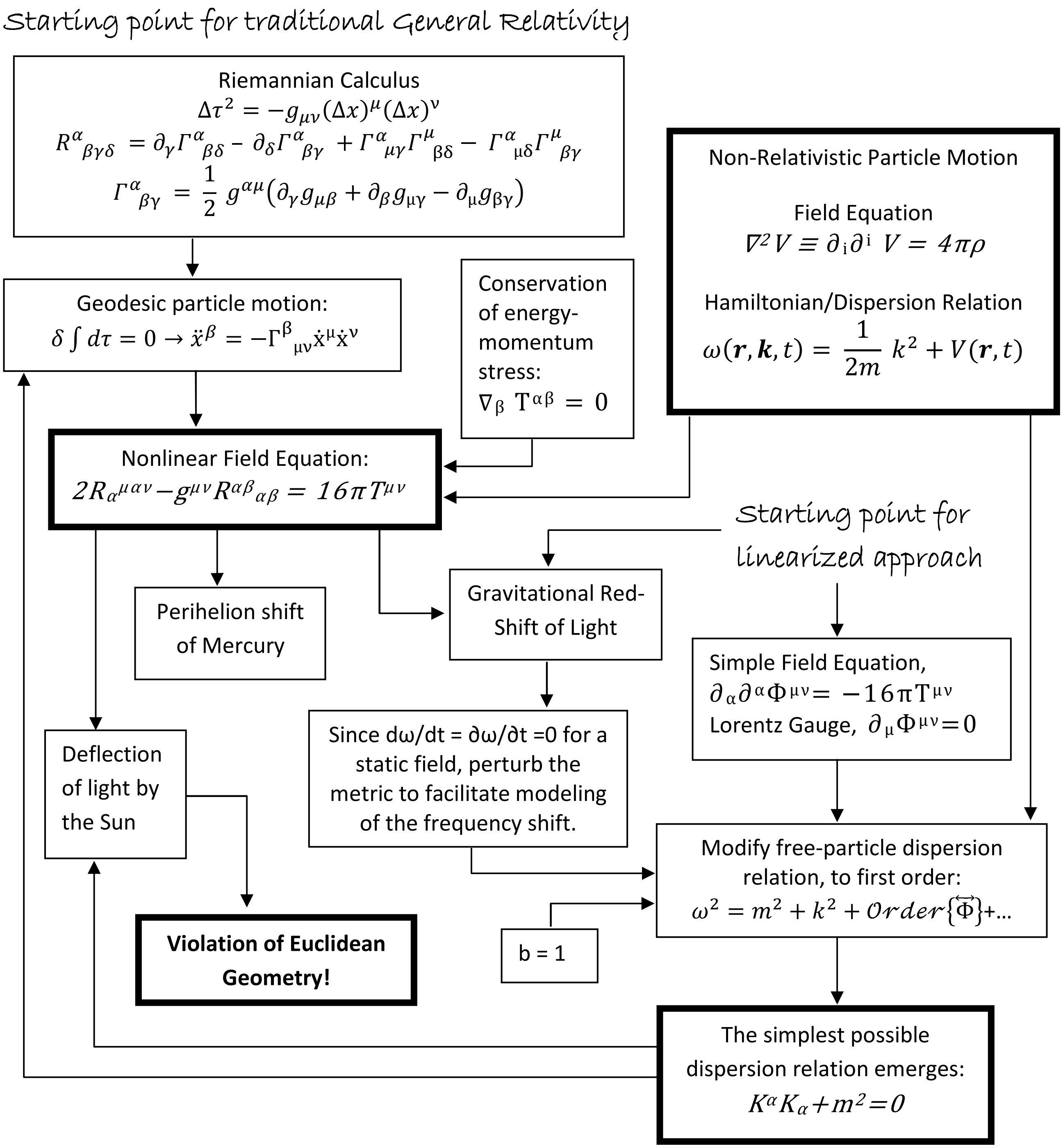

Figure 1. Concept map showing two approaches to General Relativity

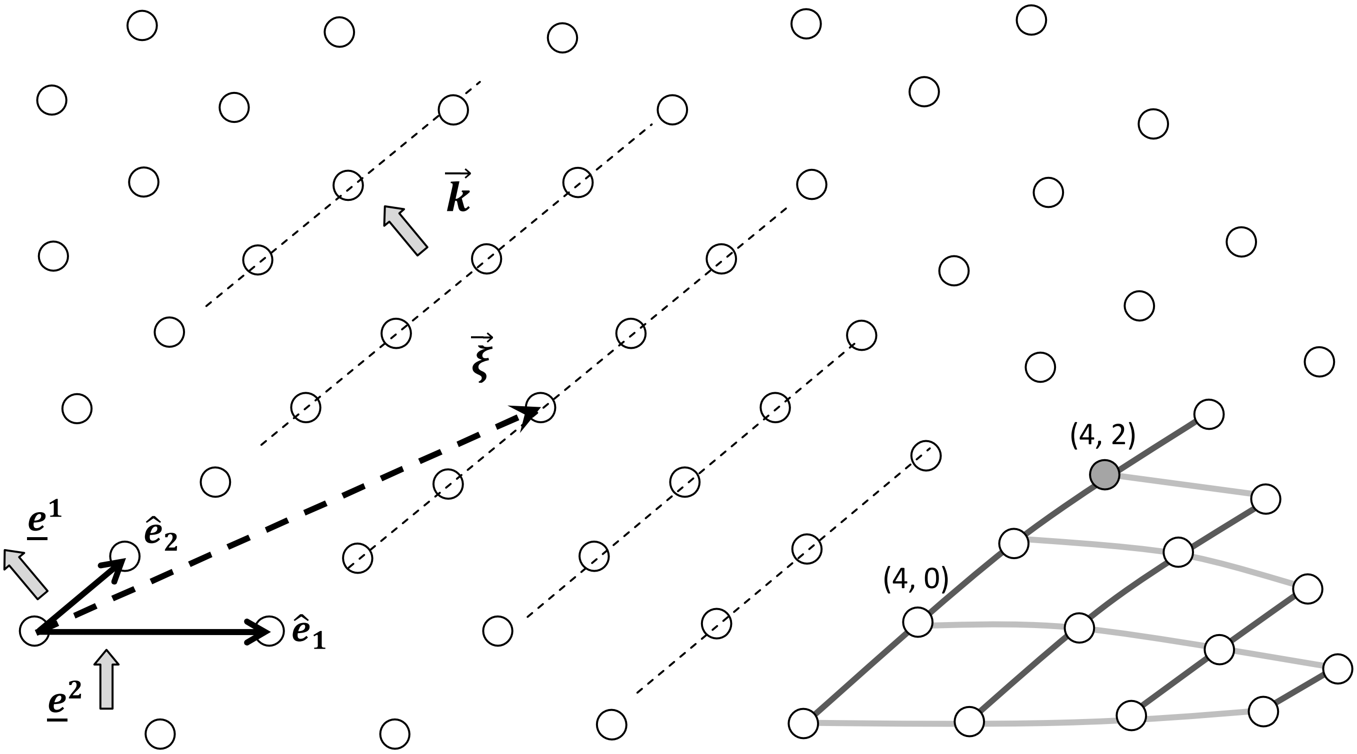

Figure 2. Circles represent atoms in inhomogeneous and anisotropic crystal lattice structure. The dashed arrow is a displacement vector, and the five parallel dashed lines represent wavefronts of localized wavepacket. Two sets of basis vectors are shown. The magnitudes of the coordinate basis vectors, $\hat{\boldsymbol{e}}_1$ and $\hat{\boldsymbol{e}}_2$, are drawn as correct, provided the inhomogeniety is sufficiently weak, and the circles represent integral values of coordinate displacements (ξ^1, ξ^2). A portion of a "net", to help label coordinates, is shown in the lower right corner.

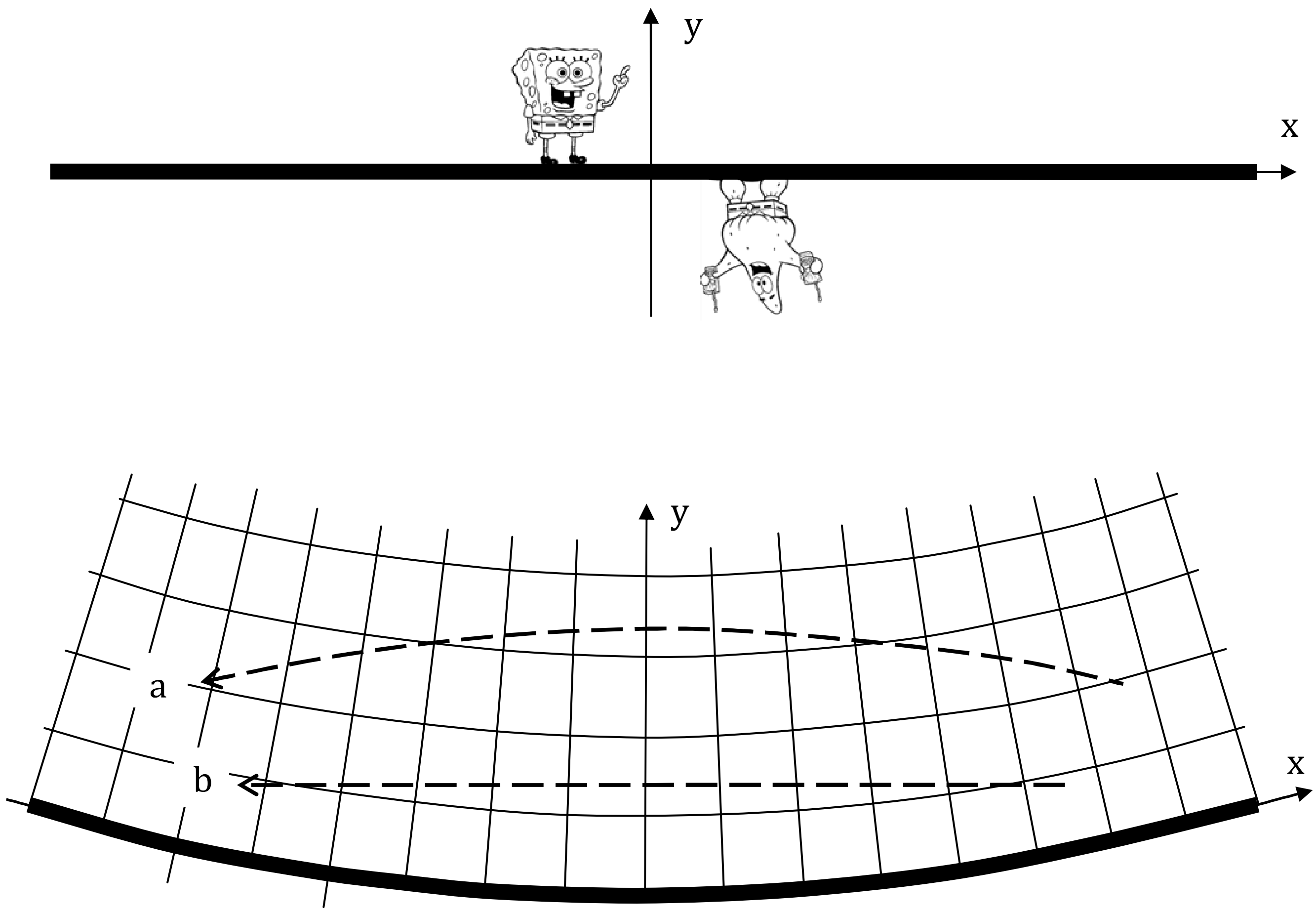

Figure 3. A planet consists of a large but very thin ultra-dense disk that generates earthlike gravity. The planet is restricted lie within the plane, y = 0, using the non-Euclidean coordinates of linearized General Relativity. The metric is such that inhabitants on each side would perceive their world to be bowl shaped if b>1. Light follows path (a) with a downward acceleration of g. There is no downward acceleration associated with straight line motion (b), even though it would exhibit a nonzero value of d^2y/dt^2.

[1] C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), 4th ed., pp. 53–4, 61–9, 85, 187-91, 415-49, 1104.

[2] H. Ohanian and R. Ruffini, *Gravitation and Spacetime*, (W. W. Norton, New York, 1994), 2nd ed., pp. 83 –5.

[3] R. D'Inverno, *Introducing Einstein's Relativity* (Clarendon Press, Oxford 1992) 271–74.

[4] M. G. Bowler, *Gravitation and Relativity* (Pergamon Press, Elmsford, NY, 1976).

[5] Ronald Adler, Maurice Bazin, and Menahem Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965)

[6] R. M. Wald , "Resource Letter TMGR-1: Teaching the mathematics of general relativity ," Am. J. Phys. **74** (6), 471-77 (2006)

[7] R. J. Cook, "Physical time and physical space in general relativity,"Am. J. Phys. **72** (2), 214-19 (2004)

[8] Steven Weinberg, "Eikonal method in magnetohydrodynamics," Phys. Rev. **126**, 1899–1909 (1962).

[9] G. Vandegrift, "The diffraction and spreading of a wavepacket," Am. J. Phys. **72** (3), 404-413 (2004)

[10] Herbert Goldstein, *Classical Mechanics*, (Addison Wesley, Reading MA, 1980), pp. 275–92, 337–9, 545–97.

[11] R. M. Jonsson, "Visualizing curved spacetime," Am. J. Phys. **73** (3) 248-60 (2005)

[12] D. Jackson*, Classical Electrodynamics* (Wiley, New York, 1962) pp. 347–86

[13] L. B. Okun, K. G. Selivanov, and V. L. Telegdi, "On the interpretation of the redshift in a static gravitational field in general relativity," Am. J. Phys. **68** (2), 115–19 (2000)

[14] R. Ferraro, "The equivalence principle and the bending of light," Am. J. Phys. **71** (2), 168-170 (2003)

[15] G. Bateman, *MHD Instabilities,* (MIT Press, Boston, 1978)

[16] S. Ichimaru, *Basic Principles of Plasma Physics,* (W. A. Benjamin, Inc., Reading, MA, 1973)

[17] N. Ashcroft and N. D. Mermin, *Solid State Physics* _Holt, Rinehart and Winston, Philadelphia, 1976), Chaps. 4 and 5.

[18] G. Vandegrift, "The maze of quantum mechanics,". Eur. J. Phys. **23** (5) 513-22 (2002)

[19] This definition is circular, with the gradient defining the dual basis, which in turn defines the gradient. However, one can define the dual basis using the conventional gradient of a scalar by appealing to the local coordinates: $\underline{\boldsymbol{e}}^{\boldsymbol{\alpha}} = \boldsymbol{i}^*\partial x^\alpha/\partial x^* + \boldsymbol{j}^*\partial x^\alpha/\partial y^* + \boldsymbol{k}^*\partial x^\alpha/\partial z^* + \boldsymbol{l}^*\partial x^\alpha/\partial t^*$, where { $\boldsymbol{i}^*$, $\boldsymbol{j}^*$, $\boldsymbol{i}^*$, $\boldsymbol{k}^*$, $\boldsymbol{l}^*$}are unit vectors in a Minkowskian space.

[20] Kuo-Ho Yang, "The physics of gauge transformations," Am. J. Phys. **73** (8), 742-51 (2005)

[21] J. C. Baez and E. F. Bunn, "The meaning of Einstein's equation," Am. J. Phys. **73** (7), 644-52 (2005)

[22] P. Jones, G. Muñoz, M. Ragsdale, and D. Singleton, "The general relativistic infinite plane," Am. J. Phys. **76** (1), 73-8 (2008)

[23] S. Deser and J. Franklin, "Time (in)dependence in general relativity," Am. J. Phys. **75** (3) 281-5 (2007)

[24] A. Harvey, E. Schucking, and E. J. Surowitz, "Redshifts and Killing vectors,"Am. J. Phys. **74** (11), 1017-24 (2006)

[25] M. McCall and D. Censor, "Relativity and mathematical tools: Waves in moving media" Am. J. Phys. **75** (12) 1134-40 (2007)