The Motion of Small Bodies in Space-time

Robert Geroch
Enrico Fermi Institute
The University of Chicago
5640 Ellis Ave., Chicago, IL, 60637, USA
geroch@uchicago.edu

James Owen Weatherall
Department of Logic and Philosophy of Science
University of California, Irvine
3151 Social Science Plaza A, Irvine, CA, 92697-5100 USA
weatherj@uci.edu

Abstract

We consider the motion of small bodies in general relativity. The key result captures a sense in which such bodies follow timelike geodesics (or, in the case of charged bodies, Lorentz-force curves). This result clarifies the relationship between approaches that model such bodies as distributions supported on a curve, and those that employ smooth fields supported in small neighborhoods of a curve. This result also applies to "bodies" constructed from wave packets of Maxwell or Klein-Gordon fields. There follows a simple and precise formulation of the optical limit for Maxwell fields.

1. Introduction

It has been generally believed, since the work of Einstein and others [1, 2], that general relativity predicts, in some sense, that a small body, free from external interaction, must move on a geodesic in space-time. But unraveling the details of that sense has turned out to be a delicate matter.[1]

Consider a physical theory that incorporates a 4-dimensional manifold M of events on which there is specified various fields, subject to a system of partial differential equations. Let this system have an initial-value formulation, i.e., be such that specification of the fields "initially" determines, by virtue of the equations, those fields subsequently. This would seem to be the minimum requirement for a (non-quantum) physical theory in space-time. Next, let there be constructed, from these fields, a material body. Then, by virtue of the initial-value formulation,

[1]For a review of the literature, see [3] and [4]. Examples of recent work in this area include [5] and [6].

every aspect of the future behavior of that body — its motion as a whole as well as its detailed internal structure — is determined already by those initial conditions. Thus, in a very broad sense, every physical theory in space-time has the feature that its equations, taken as a whole, determine the motion of bodies in that theory.

But the prediction of general relativity is supposed to be more specific: That every such body, independent of its composition and construction, moves, roughly, "along a geodesic". How does this come about? A key feature of general relativity is that all matter must couple to gravity by means of a stress-energy tensor field, T^{ab}, and that this field, by virtue of the equations on that matter, must be conserved. In effect, general relativity singles out, from the full set of equations on the matter fields, a certain subset. The idea is that this subset, which reflects (local) conservation of energy and momentum, governs the motion of the body as a whole, while the remaining matter equations govern the behavior of the internal structure of the body. It is this universal conservation of stress-energy (together with a suitable energy condition [7]) that is to lead to geodesic motion in general relativity.

Fix an exact solution of Einstein's equation, and let there be singled out, in this solution, a "body". For example, in an exact solution representing the sun-earth system, we might single out the earth. We would like to assert that such a body must be moving along a geodesic. The problem, of course, is that we are dealing with an extended body, and it is not clear what "along a geodesic" means.

This problem does not arise, for example, in Newtonian gravity, for there we have a general result: The acceleration of the center of mass of an extended body is given by the mass-weighted average gravitational field acting on that body. One strategy would be to try to extend this Newtonian result to general relativity. For example, Dixon [8] defines an "effective center of mass" trajectory for an extended body in general relativity. The acceleration of that trajectory then serves as a proxy for the acceleration of the body as a whole. But, at least in some cases, this is not a good approximation, for the acceleration of this trajectory can be strongly affected by the local matter distribution in its vicinity — e.g., by a nearby small, massive rock.

What one would like, ideally, is to produce, for any given extended body in general relativity, a number that represents "the extent to which that body, taken as a whole, fails to follow a geodesic". Then, one would like to produce a suitable upper bound on that number. Thus, this bound would describe which features an extended body must possess in order that its motion be approximately along a geodesic. But such a general bound seems well out of reach.

Here is an alternative strategy. The first step is to find a context in which the "motion of the body", as well as the issue of whether or not that motion is geodesic, make sense. That context involves a body small in size (so we know where it is), and small in mass (so its mass does not, through Einstein's equation, render the space-time metric badly behaved). Such a context is the following: Fix a space-time, (M, g_{ab}), together with a timelike curve γ in that space-time. Here, the space-time is intended to represent an idealized version of "the world with the body absent"; while the curve represents an idealized "path of the body". Thus, in the case of the earth, (M, g_{ab}) might be the Schwarzschild solution, and γ a curve orbiting the central (solar) mass. Of course, this choice is somewhat subjective, for there is no algorithm in general relativity for constructing "the same space-time, but with the body removed". The second step is to insert, into this context, some geometrical object to provide an idealized representation of the material content of the body itself. The idea is that, insofar as the actual space-time with the actual extended body resembles this idealized version, then to that extent the curve γ reflects "the motion of the actual body". Absent the sort of precise bound described above, we are reduced to such a comparison.

There are two candidates for this geometrical object. One, originating with Matthison [9] and developed by Souriau [10], Sternberg and Guillemin [12], and others, represents the body by a certain stress-energy distribution $\mathbf{T}^{ab}$ with support on γ. Provided this distribution is of order zero and has vanishing divergence, it follows that γ must be a geodesic. The other [13, 14] represents the body by a family of smooth tensor fields, T^{ab}, on the space-time. Provided there exist such T^{ab} that are conserved, satisfy an energy condition, and are supported in arbitrarily small neighborhoods of γ, it again follows that γ must be a geodesic. The distribution $\mathbf{T}^{ab}$ or tensor field T^{ab} represents a sort of idealized limiting stress-energy, i.e., the result of scaling the actual stress-energy of a body by a factor that increases as the mass goes to zero, so arranged to achieve a finite limit.

Both of these approaches, then, contemplate a limit of a body small in size and mass. They are similar in spirit, but quite different in their structure. Our goal is to connect them.

In Sect 2 we discuss the distributional approach. We show that, for a distribution $\mathbf{T}^{ab}$ supported on a timelike curve γ, conservation and an energy condition already imply that $\mathbf{T}^{ab}$ must be the stress-energy of a "point mass", i.e., a multiple of $u^a u^b \boldsymbol{\delta}_\gamma$, where u^a is the unit tangent to γ and $\boldsymbol{\delta}_\gamma$ is the delta distribution of γ; and that γ must be a geodesic. A similar result is available even if conservation is relaxed, i.e., in the presence of external forces. There is a certain freedom, in this case, to incorporate stresses, originally in $\mathbf{T}^{ab}$, into the force. Exploiting this freedom, there results, for $\mathbf{T}^{ab}$, again a point-mass stress-energy; and, for γ, a curve satisfying Newton's Law. We also consider, in Sect 2, the case of a particle carrying charge-current, subject to electromagnetic forces. It follows, quite generally, from this distributional treatment that all electromagnetic moments of the particle of order higher than dipole must vanish. The final equation of motion for such a particle is what we expect: The Lorentz force plus certain additional forces arising from the interaction between the dipole moments and the gradients of the external field.

In Sect 3, we show how these distributional stress-energies arise as limits of smooth stress-energies, in the spirit of [13, 14]. The main result, Theorem 3, is to the effect that if a collection $\mathcal{C}$ of smooth fields, T^{ab}, conserved and satisfying an energy condition, tracks sufficiently closely a timelike curve γ, then some sequence of fields from that collection actually converges, up to a factor, to the point-mass distribution; and, therefore, from conservation, that γ must be a geodesic. The key feature of this theorem is that we do *not* require that the members of $\mathcal{C}$ converge to anything at all, but only that they track, in a suitable sense, the curve γ. Then convergence — not only to some distribution, but to a specific one — follows. In this sense, then, the distributional description of particle motion reflects the strategy of [13, 14]. Theorem 3 has a simple and natural generalization to bodies carrying charge.

In section 4, we consider a class of examples: Bodies consisting of wave packets constructed from Maxwell or Klein-Gordon fields. It turns out that the results of section 3 apply to such bodies. We thus provide a simple proof that all such packets, in an appropriate limit, follow geodesics (or, in the case of charged fields, Lorentz-force curves). The well-known "optical limit" of electromagnetic waves is a special case.

Section 2. Particles

In this section, we consider the motion of a body in the limit in which that body is confined entirely to a single curve, i.e., the limit in which the "path of the body" makes sense. We must describe the body, in this limit, in terms of distributions. A few facts regarding distributions are summarized in Appendices A and B. Our purpose here is merely to introduce these distributions and describe their properties. The motivation for this treatment appears in Sect 3, in which we discuss the sense in which the present limit describes the behavior of actual, i.e., extended, bodies.

Fix, once and for all, a smooth space-time, (M, g_{ab}).

A symmetric tensor T^{ab} at a point of this space-time is said to satisfy the *dominant energy condition*[2] at that point provided: For any two timelike or null vectors, u_a, v_b, at that point, lying in the same half of the light cone, $T^{ab}u_a v_b \geq 0$. A symmetric tensor t_{ab} at that point is said to satisfy the *dual energy condition* provided $t_{ab}T^{ab} \geq 0$ for every T^{ab} satisfying the dominant energy condition, i.e., provided t_{ab} can be written as a sum of symmetrized outer products of pairs of timelike or null vectors, all lying in the same half of the light cone. Such a t_{ab} will be called *generic* in case this inequality is strict whenever T^{ab} is nonzero. The (closed) cones of tensors satisfying the dominant energy condition and the dual energy condition are duals of each other. The latter cone a proper subset of the former, and the interior of that cone consists precisely of the generic t_{ab}.

A symmetric distribution, $\mathbf{T}^{ab}$, will be said to satisfy the (dominant) *energy condition* provided $\mathbf{T}\{\textsc{t}\} \geq 0$ for every test field $\textsc{t}_{ab}$ satisfying the dual energy condition everywhere.

A key fact about distributional stress-energies is the following:

Theorem 1. Let $\mathbf{T}^{ab}$ be a symmetric distribution, satisfying the energy condition. Then $\mathbf{T}^{ab}$ is order zero.

Proof: Let $\textsc{t}^1{}_{ab}, \textsc{t}^2{}_{ab}, ...$ be a sequence of test fields, with common compact support, C^0-converging to test field $\textsc{t}_{ab}$. Fix any test field, $\textsc{s}_{ab}$, satisfying the dual energy condition and generic on the supports of $\textsc{t}_{ab}$ and the $\textsc{t}^n{}_{ab}$. Then, for every $\epsilon > 0$, both $\epsilon\textsc{s}_{ab} - \textsc{t}_{ab} + \textsc{t}^n{}_{ab}$ and $\epsilon\textsc{s}_{ab} + \textsc{t}_{ab} - \textsc{t}^n{}_{ab}$ satisfy the dual energy condition for all sufficiently large n. Applying $\mathbf{T}^{ab}$ to these two fields, we conclude: $|\mathbf{T}(\textsc{t}) - \mathbf{T}(\textsc{t}^n)| \leq \epsilon\mathbf{T}(\textsc{s})$ for all sufficiently large n.

In fact, the conclusion of Theorem 1 holds for any distribution "arising from tensors whose value, at each point, is restricted to some proper convex cone". For example, it holds also for nonnegative scalar distributions and for future-directed timelike vector distributions, as well as for symmetric tensor distributions satisfying various other energy conditions (suitably defined).

Next, let γ be a timelike curve[3] on this space-time. A particle traversing γ is represented by its stress-energy: a nonzero symmetric distribution, $\mathbf{T}^{ab}$, satisfying the energy condition and

[2] We shall use the dominant energy condition throughout. In a substantial fraction of cases, nothing is changed by replacing this by the weak, or the strong, energy condition.

[3] We shall take all curves to be smooth, connected, embedded, without endpoints and, for timelike curves, parameterized by length.

supported on γ. Set $\mathbf{f}^a = \nabla_b \mathbf{T}^{ab}$, the (four-)force (density) that drives $\mathbf{T}^{ab}$. So, for example, were there no such force, $\mathbf{f}^a = 0$, then the stress-energy would be conserved. This distribution $\mathbf{f}^a$ is also supported on γ, and is necessarily of order one, by virtue of the fact that it is expressed as the derivative of an order-zero distribution.

Fix any smooth vector field u^a on M that, at points of γ, is unit and tangent to this curve. (Everything will be independent of how this u^a is extended off γ.) We may now decompose $\mathbf{T}^{ab}$ into its spatial and temporal parts:

$$\mathbf{T}^{ab} = \boldsymbol{\mu} u^a u^b + 2u^{(a} \boldsymbol{\rho}^{b)} + \boldsymbol{\sigma}^{ab}. \tag{1}$$

Here, $\boldsymbol{\mu}, \boldsymbol{\rho}^a$ and $\boldsymbol{\sigma}^{ab} = \boldsymbol{\sigma}^{(ab)}$ are (unique) order-zero distributions, supported on γ, and having all their indices spatial (i.e., for example, $\boldsymbol{\rho}^b u_b = 0$). Here, $\boldsymbol{\mu}$ is interpreted as the mass density of the particle, $\boldsymbol{\rho}^a$ as an internal momentum density, and $\boldsymbol{\sigma}^{ab}$ as the stress density. That the distribution $\boldsymbol{\mu}$ is order zero means that this particle can manifest no higher mass multipole moments: Indeed, an n-pole mass distribution is of order n. We note that the energy condition on $\mathbf{T}^{ab}$ imposes on these three distributions a certain inequality — roughly speaking, that $\boldsymbol{\mu}$ be nonnegative and that it bound both $\boldsymbol{\rho}^a$ and $\boldsymbol{\sigma}^{ab}$.

We now turn to the force distribution $\mathbf{f}^a$. Its decomposition is more complicated than that of $\mathbf{T}^{ab}$, because $\mathbf{f}^a$ is order one rather than zero. It is convenient to introduce the following notion.

A distribution with support on γ will be called *local* (to γ) provided it annihilates every test field that vanishes on γ. For example, every order-zero distribution with support on γ — such as $\boldsymbol{\delta}_\gamma$ and the delta distribution of any one point of γ — is automatically local. For order-one distributions with support on γ, some are local and some are not. A distribution of this order is local, roughly speaking, if it "takes the derivative (of a test field to which it is applied) only in the direction along γ, and not in any orthogonal directions". Indeed, let $\boldsymbol{\alpha}^a{}_X$ be any local distribution (where here "X" represents any arrangement of indices). Then $\nabla_a \boldsymbol{\alpha}^a{}_X$ is also local if and only if $\boldsymbol{\alpha}^a{}_X$ is a multiple of the tangent, u^a, to γ. Furthermore, every local distribution is of this form, i.e., is of the form $\nabla_a(u^a \boldsymbol{\beta}_X)$ for some local $\boldsymbol{\beta}_X$. And, finally, this $\boldsymbol{\beta}_X$ is unique up to adding to it $\zeta_X \boldsymbol{\delta}_\gamma$, where ζ_X is a tensor parallel-transported along γ. In short, we may "take the integral, along γ", of any local distribution.

The usefulness of the notion of a local distribution stems from the following fact.

Theorem 2. Let $\mathbf{f}^a$ be any order-one distribution with support on γ. Then

$$\mathbf{f}^a = \boldsymbol{\alpha}^a + \nabla_b \boldsymbol{\beta}^{ab}, \tag{2}$$

where $\boldsymbol{\alpha}^a$ is a local, order-one distribution; and $\boldsymbol{\beta}^{ab}$ is an order-zero distribution, spatial in the the index b, with support on γ. Furthermore, the distributions $\boldsymbol{\alpha}^a$ and $\boldsymbol{\beta}^{ab}$ are unique.

Proof. From the fact that $\mathbf{f}^a$ is order one, we have: There exist order-zero distributions $\boldsymbol{\xi}^a$ and $\boldsymbol{\psi}^{ab}$ such that $\mathbf{f}\{\mathrm{T}\} = \boldsymbol{\xi}\{\mathrm{T}\} + \boldsymbol{\psi}\{\nabla \mathrm{T}\}$ for every test field T_a. Set $\boldsymbol{\beta}^{ab} = -q^b{}_c \boldsymbol{\psi}^{ac}$ and $\boldsymbol{\alpha}^a = \boldsymbol{\xi}^a + \nabla_c(u^c u_d \boldsymbol{\psi}^{ad})$, where $q^b{}_c = \delta^b{}_c + u^b u_c$ is the spatial projector. Eqn. (2) follows. Uniqueness is immediate.

Eqn. (2) provides a natural decomposition of the force, $\mathbf{f}^a$, on a particle into a local part (the first term on the right) and a nonlocal part (the second term). Physically, the distribution $\boldsymbol{\alpha}^a$ in (2) represents a total force acting on the particle as a whole. The distribution $\boldsymbol{\beta}^{ab}$, by contrast, can be interpreted as describing a "dipole force": a pair of equal and opposite forces, in the limit in which the magnitudes of those forces become large while at the same time the points at which those forces act become closer together. This is, for example, the force distribution produced by an electric dipole moment in a constant external electric field. Theorem 2 can be generalized to distributions of higher-order, but not, apparently, to distributions with more general support.

Substituting (1) and (2) into the force-equation, $\nabla_b \mathbf{T}^{ab} = \mathbf{f}^a$, we obtain

$$\nabla_b \boldsymbol{\sigma}^{ab} + \nabla_b(u^a \boldsymbol{\rho}^b) + \nabla_b(\boldsymbol{\rho}^a u^b) + \nabla_b(\boldsymbol{\mu} u^a u^b) = \boldsymbol{\alpha}^a + \nabla_b \boldsymbol{\beta}^{ab}. \tag{3}$$

This equation describes the response of the various components of the stress-energy to a given external force. We first note that, from uniqueness in Theorem 2, there follows

$$\boldsymbol{\sigma}^{ab} + u^a \boldsymbol{\rho}^b = \boldsymbol{\beta}^{ab}. \tag{4}$$

Eqn. (4) implies that the stress of the particle, $\boldsymbol{\sigma}^{ab}$, as well as its internal momentum, $\boldsymbol{\rho}^a$, are already completely determined by the (nonlocal part of the) force. In physical terms, the particle must react to such an external, nonlocal force by adjusting its internal stress and momentum to accommodate that force. It also follows from Eqn. (4) that the spatial projection of $\boldsymbol{\beta}$ must be symmetric:

$$\boldsymbol{\beta}^{m[b} q^{a]}{}_m = 0. \tag{5}$$

Physically, the left side of (5) represents the torque imposed on the particle by the (nonlocal part of the) external force $\mathbf{f}^a$. Eqn. (5), then, reflects the fact that a point particle is unable to absorb torque (storing it internally as (spin) angular momentum) for any such storage would violate the energy condition. Were a body to attempt to arrange itself so as to violate (5), then it would, in the limit of a point particle, be forced to adjust its spatial orientation, instantaneously, in just such a way that (5) is restored.

As an example of the above, let us consider the special case of a "free particle": $\mathbf{f}^a = 0$, and so $\mathbf{T}^{ab}$ conserved. Then (4) implies that $\boldsymbol{\sigma}^{ab} = 0$ and $\boldsymbol{\rho}^a = 0$. Thus, the stress-energy in this case is given simply by $\mathbf{T}^{ab} = \boldsymbol{\mu} u^a u^b$, i.e., is that of a "mass point". Conservation of this stress-energy yields $\nabla_a(\boldsymbol{\mu} u^a) = 0$ and $\boldsymbol{\mu} u^a \nabla_a u^b = 0$. The former implies that $\boldsymbol{\mu} = m \boldsymbol{\delta}_\gamma$, where m is a positive number, interpreted as the mass of the particle, and $\boldsymbol{\delta}_\gamma$ is the delta distribution of γ. The latter implies that the curve γ is a geodesic. Thus, we recover in this special case what we expect: a point particle, of constant mass, moving on a geodesic in space-time. An important feature of this example should be noted: We only impose on the distribution $\mathbf{T}^{ab}$ conservation, the energy condition, and that its support be on γ — but nothing about the form that $\mathbf{T}^{ab}$ must take. From only this input, the specific form $\mathbf{T}^{ab} = m u^a u^b \boldsymbol{\delta}_\gamma$ already follows.

We return now to the general case, with $\mathbf{f}^a \neq 0$. We have been thinking of $\mathbf{f}^a$ as an effective force, imposed on the particle by its external environment. This "external environment" itself consists of some additional matter — possibly of the same type as that of which the particle is composed, possibly of some different type. But when two samples of matter are in interaction with each other, there is in general no clear-cut way to decide how the matter is to be allocated between the two samples. Consider, for example, a body carrying a charge distribution, placed in a background electromagnetic field. There results a total electromagnetic field. How is the momentum stored in this total field to be allocated between the body and the background?

This freedom shows itself, in the present instance, in the ability to add, to $\mathbf{T}^{ab}$, any symmetric, order-zero distribution $\mathbf{s}^{ab}$ with support on γ, and simultaneously to add, to $\mathbf{f}^a$, the divergence of that distribution. As noted above, for an ordinary body, of finite size, there is in general no natural way to resolve this ambiguity between the body and its environment. But it turns out that, in the "particle limit", there is such a way. We choose $\mathbf{s}^{ab} = -\boldsymbol{\beta}^{ab} - u_m \boldsymbol{\beta}^{ma} u^b$. This $\mathbf{s}^{ab}$ is symmetric, by virtue of (4); and, furthermore, results in an order-zero total force. Indeed, this $\mathbf{s}^{ab}$ is the most general tensor distribution, constructed from $\boldsymbol{\beta}^{ab}$, having these properties. In physical terms, we are allocating any "dipole force" that may be acting on the particle entirely to the particle. As a result of this adjustment, we obtain, from (4), $\mathbf{T}^{ab} = \boldsymbol{\mu} u^a u^b$. In short, this adjustment converts a general $\mathbf{T}^{ab}$ (given by (1)) and $\mathbf{f}^a$ (given by (2)) into a new pair having stress-energy that of a "point particle", subject to a force that is local. This new pair is physically equivalent to the original $(\mathbf{T}^{ab}, \mathbf{f}^a)$: They describe exactly the same physical situation, but with a different choice of variables.

The force equation, $\nabla_b \mathbf{T}^{ab} = \mathbf{f}^a$, now reduces to

$$\boldsymbol{\mu} A^a = q^a{}_b \mathbf{f}^b, \tag{6}$$

$$\nabla_b(\boldsymbol{\mu} u^b) = -\mathbf{f}^b u_b, \tag{7}$$

where A^a is the acceleration of the curve γ, and $q^a{}_b$ is the spatial projector. We interpret Eqn. (6) as Newton's second law. Eqn. (7) represents conservation of mass: The particle may gain or lose mass by virtue of any time-component of the force[4]. We emphasize that, in (6)-(7), $\boldsymbol{\mu}$ is simply some non-negative, order-zero distribution supported on γ. It could, for example, involve $\boldsymbol{\delta}$-distributions of points of γ.

There is a certain sense in which Eqns. (6)-(7) manifest an initial-value formulation. Imagine, for a moment, that some force-distribution, $\mathbf{f}^a$, were specified on the manifold M, once and for all. We wish to insert a particle into this environment. To this end, we choose a point p of M (the initial position of the particle), a unit timelike vector u^a at p (the initial 4-velocity of the particle), and a number $m > 0$ (the initial mass of the particle). Then: The evolution of the point p is determined by u^a; the evolution of u^a is determined (via A^a) by (6); and the evolution of m is determined by (7). Thus, the rates of change of these three objects, at each point, are determined by the values of these objects at that point. In other words, we expect to be able to determine the future evolution of the particle.

But there is a problem with such a formulation, having to do with the nature of the force itself. Eqns. (6) and (7) each assert that one distribution (on M) is equal to another. The distributions on the left (arising from the particle) have support on γ, and so, therefore, must the distribution $\mathbf{f}^a$. Thus, it makes no sense to "specify the force, once and for all, as a distribution on M". Instead, we must specify $\mathbf{f}^a$ as a distribution *with support on* γ. But we cannot do this, within the context of an initial-value formulation, because we don't know ahead of time

[4]It is tempting at this point to make a further adjustment between $\mathbf{T}^{ab}$ and $\mathbf{f}^a$, so as to achieve $\mathbf{f}^a u_a = 0$. Such adjustments exist, but, unfortunately, there does not appear to be any single, natural one.

where the curve γ will be! What often happens in specific examples (as we shall see shortly) is that the particle itself is involved in the determination of $\mathbf{f}^a$, in this manner achieving an $\mathbf{f}^a$ supported on γ.

There is a curious interaction between (6) and (7). In these equations, the distribution $\boldsymbol{\mu}$ is order zero, while the distribution $\mathbf{f}^a$ is permitted to be order one. Now, it follows, from the fact that $\boldsymbol{\mu}$ in (6) is order zero, that the spatial force, $q^a{}_b\mathbf{f}^b$, must be zero order. It is certainly possible to write down a total force, $\mathbf{f}^a$, such that its spatial component is order zero, but its temporal component is genuinely of order one. But actually to achieve such a total force, physically, would seem to be a rather delicate business: It would mean that the environment, while transferring mass to the particle in a manner that is genuinely order one, at the same time avoids transferring momentum to that same order. One would expect that the slightest error in the transfer process would result in an order-one spatial force, and, as a consequence, in a curve γ that fails to be smooth. Arguably, such delicate adjustments are unphysical. On this basis, then, let us now demand that the temporal component of the force — and hence the entire $\mathbf{f}^a$ — be order zero. But now we can repeat the same argument. Since now the right side of (7) is order zero, μ must be of the form $m\boldsymbol{\delta}_\gamma$, where m is a locally integrable function on γ. But now (6) implies that the spatial component of the force have the same character. On the same physical grounds as above, we may demand that all of $\mathbf{f}^a$ be a locally integrable vector times $\boldsymbol{\delta}_\gamma$. Repeating this same argument, over and over, we finally conclude: Each of $\boldsymbol{\mu}$ and $\mathbf{f}^a$ must be of the form of a smooth field along γ multiplied by $\boldsymbol{\delta}_\gamma$. In this case, (6)-(7) becomes a simple system of ordinary differential equations along the curve. We emphasize that the above is merely a rough plausibility argument.

We remark that it is easy to write down explicitly the general stress-energy distribution for a (not necessarily free) particle. Fix any timelike curve γ, and any non-negative, order-zero distribution $\boldsymbol{\mu}$ supported on γ. Then set $\mathbf{T}^{ab} = \boldsymbol{\mu} u^a u^b$ and $\mathbf{f}^a = \nabla_b(\boldsymbol{\mu} u^a u^b)$.

We have already noted a special case of (6)-(7): that of a "free particle" (i.e., that of $\mathbf{f}^a = 0$). Then $\boldsymbol{\mu} = m\boldsymbol{\delta}_\gamma$, where $m > 0$ is a number; and the curve γ is a geodesic. Another special case of interest is that of a charged particle. Let there be given a fixed, smooth, antisymmetric tensor field F_{ab} on M, the Maxwell field. We think of this F_{ab} as generated by some external charge-current distribution. Further, let our particle manifest its own charge-current distribution $\mathbf{J}^a$. Then $\mathbf{J}^a$ must be conserved, $\nabla_a\mathbf{J}^a = 0$, and must have support on γ. The force $\mathbf{f}^a$ is then the Lorentz force, $\mathbf{f}^a = F^a{}_b\mathbf{J}^b$, that F_{ab} imposes on $\mathbf{J}^a$. In order to guarantee that this force be order one (which is necessary, since $\mathbf{f}^a$ equals the divergence of the zero-order distribution $\mathbf{T}^{ab}$) we demand that this charge-current $\mathbf{J}^a$ also be order one. Note that there is no "self-force" here, i.e., no term involving the interaction of $\mathbf{J}^a$ with the electromagnetic field it produces: Such an interaction term would vanish in the present limit.

We consider first a special case: We demand that $\mathbf{J}^a$ actually be order zero. It then follows (from uniqueness in Theorem 2) that $\mathbf{J}^a = eu^a\boldsymbol{\delta}_\gamma$, where e is some number, which we interpret as the total charge of the particle. Now substitute the resulting Lorentz force into (6)-(7). It follows from (7) that $\boldsymbol{\mu} = m\boldsymbol{\delta}_\gamma$, where m is, again, a positive number representing the mass of the particle. Finally, (6) implies that the curve γ is a Lorentz-force curve with mass m and charge e. In short, we recover in this special case the standard equation of motion for a charged point particle.

We turn next to the case of a particle carrying a general charge distribution — that in which the charge-current $\mathbf{J}^a$ is fully of order 1. The most general conserved, order-one vector distribution $\mathbf{J}^a$ with support on γ is given by

$$\mathbf{J}^a = eu^a \boldsymbol{\delta}_\gamma + \nabla_b(\boldsymbol{\tau}^{ab}), \tag{8}$$

where e is a number, and $\boldsymbol{\tau}^{ab}$ is some antisymmetric, order-zero distribution (not necessarily spatial) with support on γ. To see this, first identify e by applying $\mathbf{J}$ to test fields given by the gradient of a function that is constant in a neighborhood of an initial, and also of a final, segment of γ; and then use the method of Theorem 2. Again, we interpret e as the total electric charge of the particle. We interpret $\boldsymbol{\tau}^{ab}$ as describing the electric and magnetic dipole moments of the particle:

$$\boldsymbol{\tau}^{ab} = 2u^{[a}\boldsymbol{\xi}_E{}^{b]} + 1/2\ \epsilon^{ab}{}_{mn}\boldsymbol{\xi}_B{}^m u^n. \tag{9}$$

Here, the moments, represented by $\boldsymbol{\xi}_E{}^a$ and $\boldsymbol{\xi}_B{}^a$, are order-zero spatial distributions with support on γ. Note that the charge-current $\mathbf{J}^a$ can manifest at most dipole — but no higher — moments. The physical reason for this is the following. Suppose that the body tried to so arrange its charges to form, e.g., a nonzero electric quadrupole moment. There would result from this arrangement large electric forces between those charges, which would then require, in order to hold those charges in place, large stresses. But large stresses require, by the energy condition, a large mass density. If we now attempt to take the particle limit, retaining a nonzero electric quadrupole moment, we end up with an infinite limiting value for the particle's mass. In short, higher electromagnetic multipole moments are, in the end, excluded by the requirement that the particle be described by a well-defined stress-energy distribution satisfying the energy condition.

Now let the force on our particle be the Lorentz force, $F^a{}_m\mathbf{J}^m$, with the charge-current given by (8), and decompose this force as in (2). Then the condition (5) becomes

$$q^{[a}{}_m q^{b]}{}_n(F^m{}_p\boldsymbol{\tau}^{pn}) = 0. \tag{10}$$

To interpret this equation physically, decompose the electromagnetic field into its electric and magnetic parts: $F_{ab} = 2E_{[a}u_{b]} + 1/2\ \epsilon_{abmn}B^m u^n$. Substituting this, and (9), we obtain, for the left side of (10), $E^{[a}\boldsymbol{\xi}_E{}^{b]} + B^{[a}\boldsymbol{\xi}_B{}^{b]}$. This will be recognized as the physical torque on electric and magnetic dipole moments placed in an external electromagnetic field. As a general rule, a body has the freedom to choose its dipole moments at will as it traverses γ. However, in the present context — the limit of a point particle — it is necessary that these moments be so chosen that (10) holds. Any attempt to do otherwise will result in a net torque on the particle, which will quickly rotate it so as to restore (10).

For the net force on this particle, after incorporating any nonlocal contributions to the force into the stress-energy, as described earlier, we obtain:

$$\mathbf{f}^a = eF^a{}_m u^m\boldsymbol{\delta}_\gamma + \boldsymbol{\tau}^{mn}\nabla_m F_n{}^a + \nabla_b[u^b(-F^a{}_n\boldsymbol{\tau}^{nm}u_m + u_n F^n{}_m\boldsymbol{\tau}^{mc}q^a{}_c)]. \tag{11}$$

The right side of (11) has a simple physical interpretation. The first term is the ordinary Lorentz force on a point charge e. The second term is the force on electric and magnetic dipole moments placed in an external field gradient. The third term is the time-derivative of a certain vector algebraic in the moments and external field. This vector plays the role of an effective energy-momentum arising from the interaction between the moments and the external field[5]. Eqn. (11), then, requires that any change in this interaction energy-momentum with time be reflected in a net force on the particle as a whole. Note that this total force, given by (11), is indeed local, and of order one.

In the above, $\boldsymbol{\tau}^{ab}$, which represents the particle's dipole moments, began as an arbitary order-zero distribution supported on γ. We then found that $\boldsymbol{\tau}^{ab}$ is not so arbitrary, for it must satisfy the condition, (10) —that the torque on the particle be zero. There then follows a net total force on the particle, given by (11). But, as it turns out, there is a further condition that must be imposed on the distributon $\boldsymbol{\tau}^{ab}$: It must be such that the force $\mathbf{f}^a$ is consistent, via (6), with a smooth curve γ. This condition is somewhat complicated, for the distribution $\boldsymbol{\mu}$ on the left side of (6) itself undergoes evolution, via (7). It turns out, however, that there is at least one simple way to achieve it: Let the distribution $\boldsymbol{\tau}^{ab}$ be given by a smooth tensor along γ, times $\boldsymbol{\delta}_\gamma$. Then $\mathbf{f}^a$, given by (11), also takes the form of a smooth vector times $\boldsymbol{\delta}_\gamma$. It further follows, from (7), that the distribution $\boldsymbol{\mu}$ must be some smooth function m along γ, times $\boldsymbol{\delta}_\gamma$. But now the $\boldsymbol{\delta}_\gamma$'s occur as universal factors in (6)-(7), and so may be cancelled out. Thus, we are left with a simple set of ordinary differential equations along the curve γ.

Throughout this section, we have been dealing with a particle moving along a timelike curve. We now consider the null case. Thus, let γ be a null curve, with tangent vector l^a. Let $\mathbf{T}^{ab}$ be a symmetric distribution, with support on γ, satisfying the energy condition. As before, this $\mathbf{T}^{ab}$ must be order-zero. For the case of a null curve, however, we can no longer decompose tensors into their spatial and temporal components, and so we cannot write $\mathbf{T}^{ab}$ as in (1). As before, the force driving this stress-energy is the order-one distribution given by $\mathbf{f}^a = \nabla_b \mathbf{T}^{ab}$.

Consider first the case of a free particle, $\mathbf{f}^a = 0$. Then conservation yields that γ is a geodesic, and also that $\mathbf{T}^{ab} = \boldsymbol{\mu} l^a l^b$, for some non-negative, order-zero distribution $\boldsymbol{\mu}$. Let us now choose an affine parameter for this geodesic, and let l^a be the corresponding affine tangent vector. Then conservation further implies that $\boldsymbol{\mu} = m\boldsymbol{\delta}_\gamma$, where $m > 0$ is a number and $\boldsymbol{\delta}_\gamma$ is the delta distribution on γ arising from this affine parameterization. Note that this number m cannot be interpreted as the "mass", for m scales under a change in the choice of affine parameter.

Finally, consider the case of a particle (traveling on a null curve) subject to a general force. The notion of a distribution local to γ still makes sense, for γ null. We still have the formula (2) for $\mathbf{f}^a$, but now we cannot require that $\boldsymbol{\beta}^{ab}$ be "spatial" in index b. So, uniqueness fails: We have the freedom to add to $\boldsymbol{\beta}^{ab}$ any distribution of the form $\boldsymbol{\zeta}^a l^b$, and to $\boldsymbol{\alpha}^a$ the (local) distribution $-\nabla_b(\boldsymbol{\zeta}^a l^b)$, where $\boldsymbol{\zeta}^a$ is any order-zero distribution supported on γ. We can still carry out the adjustment, as in the timelike case, to achieve $\mathbf{T}^{ab} = \boldsymbol{\mu} l^a l^b$ and $\mathbf{f}^a$ order zero. But now that adjustment is not unique: There remains the freedom to move a portion of the distribution $\boldsymbol{\mu}$ into $\mathbf{f}^a$. The force law, $\nabla_b(\boldsymbol{\mu} l^b) l^a + \boldsymbol{\mu} A^a = \mathbf{f}^a$, now requires that $\mathbf{f}^a l_a = 0$. Non-geodesic null curves are permitted.

---

[5]The issue of what should be called the "interaction energy-momentum" is somewhat tricky. An obvious strategy would be to introduce a (distributonal) solution, $\tilde{\mathbf{F}}^{ab}$, of Maxwell's equations for charge-current $\mathbf{J}^a$. Then take for the "effective force" the divergence of the cross-term in the stress-energy of the total electromagnetic field $\tilde{\mathbf{F}}^{ab} + F^{ab}$. Unfortunately, this strategy will not work here, for the resulting "force" will not be local to γ, and furthermore will depend on which solution $\tilde{\mathbf{F}}^{ab}$ is chosen. It is perhaps surprising, then, that there turns out to be any natural candidate at all for a total electromagnetic force.

3. Extended Bodies

In Sect 2, we discussed the motion of a particle — an idealized body, whose path is represented by a curve in space-time and whose matter is represented by a certain stress-energy distribution having support on that curve. This treatment turns out to be remarkably simple. We can write out, explicitly and generally, the distributions representing the matter as well as any forces that might be acting on the particle. We then determine explicitly the effect of those forces on the motion and composition of the particle.

But this of course is an idealization: Actual physical bodies have finite size. Our goal in this section is to understand the sense in which actual bodies are represented by these idealizations. Fix a space-time, satisfying Einstein's equation, in which there has been identified a "body". The general strategy, as discussed in Sect. 1, is the following. First write down an idealized space-time and body, $(M, g_{ab}, \gamma, \mathbf{T}^{ab})$, which in some sense resembles the original system. We now wish to compare the actual extended body with its idealization. To this end, we introduce some space-times, also satisfying Einstein's equation, that are intermediate between these two.

To fix ideas, let us consider first the simplest and most manageable choice of "intermediate bodies". First, introduce a family of bodies whose masses go to zero, while retaining the extended character of the original body. The space-time metrics for these bodies approach some fixed background metric, g_{ab}, on the manifold M; while their stress-energies go to zero, giving rise to a linearized field[6], T^{ab}, on M, defined only up to an overall factor. This T^{ab} inherits the energy condition from its predecessors. Next, allow the sizes of the bodies to go to zero. Thus, we end up with a collection of fields, T^{ab}, on a fixed space-time, which collapse down onto some timelike curve γ.

We must introduce a suitable sense of this "collapsing down". To this end, fix a space-time, (M, g_{ab}), and a curve γ in this space-time. Fix also some collection $\mathcal{C}$ of symmetric tensor fields T^{ab} on M, each of which satisfies the (dominant) energy condition. We will say that this collection *tracks* γ provided: Given any test field x_{ab}, satisfying the dual energy condition in a neighborhood of γ and generic[7] at some point of γ, there exists an element T^{ab} in the collection $\mathcal{C}$ such that $\mathbf{T}\{\mathrm{x}\} > 0$. This is a key definition. Note that for it we impose the energy condition (which plays a crucial role), but not conservation.

The idea of this definition is the following. Let a collection $\mathcal{C}$ track a timelike curve γ. Consider a test field x_{ab} that satisfies the dual energy condition in a narrow neighborhood of γ, but then, just outside that neighborhood, goes quickly to a very large negative multiple of a field satisfying the dual energy condition. By tracking, there must be a T^{ab} in $\mathcal{C}$ with $\mathbf{T}\{\mathrm{x}\} > 0$. In the integral that comprises $\mathbf{T}\{\mathrm{x}\}$, that neighborhood will contribute positively; and the region outside negatively. Thus, the bulk of T^{ab} must lie within this narrow neighborhood. But tracking requires that the collection $\mathcal{C}$ contain such a T^{ab} for *every* such test field x_{ab}. In short, tracking means that $\mathcal{C}$ includes fields T^{ab} the vast majority of whose matter clings, as closely as we wish and for as long as we wish, to γ.

---

[6]In more detail, we imagine a one-parameter family, $g(\lambda)_{ab}$, of metrics on M, jointly smooth in λ and point of M, satisfying Einstein's equation with stress-energy tensors $T(\lambda)^{ab}$. Then set $g_{ab} = g(0)_{ab}$ and $T^{ab} = (d/d\lambda)T(\lambda)^{ab}|_{\lambda=0}$.

[7]Recall, from Sect 2, that a tensor x_{ab}, is said to satisfy the dual energy condition at a point provided $T^{ab}x_{ab} \geq 0$ for every T^{ab} at that point satisfying the (dominant) energy condition; and to be *generic* at that point provided this inequality is strict whenever $T^{ab} \neq 0$.

Note that a given collection $\mathcal{C}$ can track more than one timelike curve. For example, the collection $\mathcal{C}$ of *all* T^{ab} satisfying the energy condition in any space-time tracks every timelike curve in that space-time. Generally speaking, modifying the T^{ab} far from a curve γ does not affect whether or not that collection tracks γ. Thus, a collection $\mathcal{C}$ could track a timelike curve γ even though *every* element of this collection includes a large amount of matter far from γ—so long as that extraneous matter manages to change its location in M, in a suitable manner (i.e., so as to avoid, eventually, every fixed test field), as we go through the various elements of $\mathcal{C}$.

We consider first the case of "free bodies", i.e., those described by stress-energies that are conserved. Here is the key theorem:

Theorem 3. Let (M, g_{ab}) be a space-time, γ a timelike curve therein, and $\mathcal{C}$ a collection of fields T^{ab}, each satisfying the (dominant) energy condition, that tracks γ. Let each of these fields be conserved. Then there exists a sequence, $\overset{1}{T}{}^{ab}, \overset{2}{T}{}^{ab}, \cdots$, each a positive multiple of some element of $\mathcal{C}$, that converges, in the sense of distributions, to $u^a u^b \boldsymbol{\delta}_\gamma$.

Proof. First, choose a test field X^0_{ab} that satisfies the dual energy condition everywhere and is generic at some point of γ, normalized by $(uu\boldsymbol{\delta}_\gamma)\{\mathrm{X}^0\} = 1$. Second, choose a sequence, $\overset{n}{\mathrm{V}}_a$, of test fields such that, setting $\overset{n}{\mathrm{X}}_{ab} = \mathrm{X}^0_{ab}/n + \nabla_{(a}\overset{n}{\mathrm{V}}_{b)}$, each $\overset{n}{\mathrm{X}}_{ab}$ satisfies the dual energy condition in a neighborhood of γ, and is generic on some segment γ_n of γ, where these segments are increasing and have union all of γ. [To do this, first fix any extension, u^a, of the tangent to γ to a unit timelike vector field, and then set $\overset{n}{\mathrm{V}}_a$ a function times u_a, where this function is so chosen to achieve the required properties.] Third, choose a sequence $\overset{1}{\mathrm{Y}}_{ab}, \overset{2}{\mathrm{Y}}_{ab}, \cdots$ of test fields, each vanishing on γ, such that i) each $\overset{n}{\mathrm{X}}_{ab} - \overset{n}{\mathrm{Y}}_{ab}$ satisfies the dual energy condition in a neighborhood of γ; and ii) for every test field M_{ab} that vanishes on γ, $\overset{n}{\mathrm{X}}_{ab} + \overset{n}{\mathrm{Y}}_{ab} - \mathrm{M}_{ab}$ satisfies the dual energy condition everywhere, for all sufficiently large n. [Choose each $\overset{n}{\mathrm{Y}}_{ab}$ to satisfy the dual energy condition and be generic wherever it is nonzero. It rises quickly off the segment γ_n, and then remains large in some region away from γ. As $n \to \infty$, the rate of rise, the size of that region, and the values of $\overset{n}{\mathrm{Y}}_{ab}$ in that region all increase.] Finally, for each n choose, by tracking, $\overset{n}{T}{}^{ab}$, a multiple of an element of $\mathcal{C}$, such that $\overset{n}{\mathbf{T}}\{\overset{n}{\mathrm{X}} - \overset{n}{\mathrm{Y}}\} > 0$, normalized by $\overset{n}{\mathbf{T}}\{\mathrm{X}^0\} = 1$.

Now let P_{ab} be any symmetric test field. Choose test vector field W_a such that $\mathrm{M}_{ab} = \mathrm{P}_{ab} - (uu\boldsymbol{\delta}_\gamma)\{\mathrm{P}\}\mathrm{X}^0_{ab} - \nabla_{(a}\mathrm{W}_{b)}$ vanishes on γ. [Here, we make use of the following fact: For $\mathrm{Z}_{ab} = \mathrm{Z}_{(ab)}$ any test field satisfying $(uu\boldsymbol{\delta}_\gamma)\{\mathrm{Z}\} = 0$, there exists a test field W_a such that $\mathrm{Z}_{ab} - \nabla_{(a}\mathrm{W}_{b)}$ vanishes on γ.] We now have, for all sufficiently large n,

$$|(\overset{n}{\mathbf{T}} - uu\boldsymbol{\delta}_\gamma)\{\mathrm{P}\}| = |\overset{n}{\mathbf{T}}\{\mathrm{M}\}| \leq \overset{n}{\mathbf{T}}\{\overset{n}{\mathrm{X}} + \overset{n}{\mathrm{Y}}\} \leq \overset{n}{\mathbf{T}}\{2\overset{n}{\mathrm{X}}\} = 2/n. \tag{12}$$

The first step follows from the definition of M, the normalization of $\overset{n}{T}$, and conservation; the second, for all sufficiently large n, from the defining property of the $\overset{n}{\mathrm{Y}}$; the third, from the defining property of $\overset{n}{T}$; and the fourth, from the defining property of $\overset{n}{\mathrm{X}}$, the normalization of $\overset{n}{T}$, and conservation. The result follows.

Theorem 3 asserts, in short, that any family of conserved stress-energies that "collapse down" onto γ, in a suitable sense, necessarily includes a sequence that converges to a certain distribution — that of a "point particle" — supported on γ. Note an important feature of the theorem. We impose on the family of T^{ab} only conditions reflecting the locations and the sizes of the bodies they represent, but no conditions on the form that T^{ab} takes; nor any on its limiting behavior. Yet, we conclude from this that some sequence from this family must converge, in a suitable sense, to *some* distribution, and, additionally, the specific form of that limiting distribution. It is easy to show from this theorem that, if a collection $\mathcal{C}$ contains a sequence that, possibly after rescaling, converges to *some* nonzero distribution supported on γ, then $\mathcal{C}$ tracks γ; and furthermore that that distribution is, up to a factor, precisely $u^a u^b \boldsymbol{\delta}_\gamma$. In short, $u^a u^b \boldsymbol{\delta}_\gamma$ is the unique distribution that arises, under conservation, from tracking. Note also that $\mathcal{C}$ is presented as merely an unordered (possibly uncountable) collection of stress-energies, with no hint as to which of its elements are close to the final distribution. The actual converging sequence is generated by the theorem.

Here, then, is a sense in which "small, free bodies in general relativity traverse geodesics". First fix the curve γ. Then demand that this curve "be followed by such bodies", in the sense that there is some collection $\mathcal{C}$ of conserved T^{ab} fields, each satisfying the energy condition, that tracks γ. Now apply Theorem 3. Since all the T^{ab} in $\mathcal{C}$ are conserved, so must be the limiting distribution, $u^a u^b \boldsymbol{\delta}_\gamma$. But, as we saw in Sect 2, conservation of $u^a u^b \boldsymbol{\delta}_\gamma$ implies that γ is a geodesic.

Here is an example of an application of the theorem. Fix a timelike curve γ in a space-time. Suppose that the following condition were satisfied: Given any compact neighborhood C of any point of γ, and any neighborhood U of γ, there exists a symmetric T^{ab}, conserved and satisfying the energy condition, that is nonzero somewhere in $C \cap U$ and vanishes in $C - U$. This condition means, in other words, that, locally, there exist conserved stress-energies that satisfy the energy condition and are confined arbitrarily closely to γ. It is immediate that the collection $\mathcal{C}$ of all the T^{ab} generated in this way tracks γ. Therefore, by Theorem 3, the curve γ must be a geodesic. This is essentially the result of [13]. (To make this comparison more transparent, we have replaced the condition on the T^{ab} in [13] by a "local" version.) But there is a significant difference between Theorem 3 and [13]: The former, but not the latter, is applicable to a collection of bodies even if every body in that collection manifests some (but not too much) matter well outside of γ. This is a useful feature, for, as we shall see in the next section, it allows us to apply the theorem to certain wave packets.

We remark that, if these $\overset{n}{T}{}^{ab}$ are expressed in terms of various matter fields, then, even though this sequence converges to a distribution, it need not be true in general that those matter fields converge to anything at all — distributional or otherwise. Indeed, the collection $\mathcal{C}$ could encompass some (idealized) stars, some rocky planets, some pieces of wood, etc.

Consider any collection of bodies, each satisfying the energy condition, that collapses down to a curve, in the sense of Theorem 3. Then, according to that theorem, the final limit must be a particle with zero spin (per unit mass). Thus, if one wishes treat "spinning particles" within this framework, then such particles must arise either i) from matter violating the energy condition, or ii) from a limit different from that envisioned in Theorem 3. Neither of these strategies appears attractive. Compare, [11].

Theorem 3 suffers from an apparent defect: It is global on the curve γ, in the sense that its hypothesis requires the existence of appropriate bodies along the entirety of this curve. Suppose, for example, that we wished to determine whether the earth, at the present epoch, travels, approximately, on a geodesic. In order to apply Theorem 3, we must introduce a timelike curve γ to represent the earth for all time; and then assert the existence of T^{ab} that track that curve, both currently and in the distant past and future. How are we to know whether this is possible (or even what γ will be) given only the earth at the present epoch? This defect, however, is easy to remedy. Fix any finite segment of γ. Then choose an open neighborhood of that segment, regard that neighborhood as a space-time in its own right, and apply in that space-time Theorem 3. We thus conclude (having only imposed conditions local to that segment) that that segment of γ must be a geodesic.

The discussion above has been for the case of "free" bodies, i.e., those represented by a conserved stress-energy. We now consider the case of bodies that interact with their environment.

The simplest case is that of a body carrying charge. Again, we imagine a process in which, first, the mass (and charge density) go to zero, maintaining the extended character of the body; and thereafter the geometrical size goes to zero. Thus, we end up with a fixed space-time, (M, g_{ab}), and a fixed timelike curve γ on that space-time. On this space-time, there is specified a fixed background electromagnetic field, F_{ab}, arising from whatever charge-current distribution is present in the environment. The body itself is described by a pair of fields, (T^{ab}, J^a), where this pair is defined only up to overall scaling, i.e., up to multiplying both fields by the same positive factor. The stress-energy T^{ab} must satisfy the energy condition and the force-law, $\nabla_b T^{ab} = F^a{}_b J^b$; and the charge-current must satisfy conservation, $\nabla_a J^a = 0$.

Now consider a collection, $\mathcal{C}$, of such pairs. Suppose that the $T's$ of this collection track γ, as described above. We would like to apply Theorem 3 to this situation. The problem, of course, is that conservation of the T^{ab}, which was used in Theorem 3, is now replaced by $\nabla_b T^{ab} = F^a{}_b J^b$. Clearly, we need to exert some control over the right side of this equation, i.e., over the charge-current J^b. Indeed, a general charge-current can manifest electric and magnetic dipole moments, and, as we saw in Sect 2, these moments can affect the motion of a body. Furthermore, there is no guarantee that, as the family of bodies collapses down onto γ, the dipole moments of these bodies will converge to anything at all. We therefore proceed as follows.

Fix, on a space-time, a pair of fields, (T^{ab}, J^a), the former satisfying the energy condition. We say that a nonnegative number κ is a bound on the charge-mass ratio for this pair provided: For any unit timelike vector t^a at any point of M,

$$|J^a t_a| \leq \kappa T^{ab} t_a t_b. \tag{13}$$

In physical terms, this means that, according to any observer located anywhere in this space-time, the ratio between the locally measured charge and mass densities of the material represented by (T^{ab}, J^a) is bounded by the number κ.

This appears to be a reasonable condition to impose on matter. It holds, for example, for any material composed (in a suitable sense) of electrons, neutrons and protons. It also holds for a charged fluid, as well as a charged stressed solid, under a suitable additional condition on the function of state[8]. Further, if the condition above holds for two types of matter, then it

[8]This condition, in the case of a charged fluid, is the following. Such a fluid is described by a 2-manifold

holds when both types are present (and allowed to overlap) in space-time. The κ-value for this combination is given by the greater of the values for the individual matter types. The condition of bounded charge-mass ratio is easily generalized to the case in which T^{ab} and J^a are both distributions. For distributions supported on a curve, as considered in Sect 2, this condition requires that the dipole and higher moments of J^a all vanish. Finally, we remark that, if the number κ bounds the charge-mass ratio for (T^{ab}, J^a), then we further have

$$|J^a J^b(g_{ab} + 2t_a t_b)|^{1/2} \leq 3\kappa T^{ab} t_a t_b \tag{14}$$

for every unit timelike t^a. In other words, we have also that the current density of the material is bounded by the mass density. Eqn. (14) also holds with the tensor $t_a t_b$ on the right replaced by a suitable multiple of any tensor satisfying the generic dual energy condition.

It turns out that the bound described above is just what is necessary to generalize Theorem 3 to charged bodies.

Theorem 4. Let (M, g_{ab}) be a space-time, F_{ab} an antisymmetric tensor field on M, and γ a timelike curve. Let $\mathcal{C}$ be a collection of pairs, (T^{ab}, J^a), of tensor fields on M, where T^{ab} satisfies the energy condition, such that each satisfies $\nabla_b T^{ab} = F^a{}_b J^b$ and $\nabla_a J^a = 0$; and each has charge-mass ratio bounded by κ, where $\kappa \geq 0$ is some fixed number. Let this collection $\mathcal{C}$ track γ. Then there exists a number κ' satisfying $|\kappa'| \leq \kappa$, along with a sequence of pairs, $(\overset{n}{T}{}^{ab}, \overset{n}{J}{}^a)$, each a multiple of some element of $\mathcal{C}$, that converges to $(u^a u^b \boldsymbol{\delta}_\gamma, \kappa' u^a \boldsymbol{\delta}_\gamma)$.

The proof consists, first, of repeating the proof of Theorem 3, including, and suitably bounding, the additional terms arising from the electromagnetic interaction. Conservation of the $\overset{n}{T}{}^{ab}$ was used at two points in that proof: In the first and fourth steps of Eqn. (12). These two steps now give rise, in (12), to additional terms $-(\nabla_b \overset{n}{\mathbf{T}}{}^{ab})\{\mathrm{w}\}$ and $-(\nabla_b \overset{n}{\mathbf{T}}{}^{ab})\{\overset{n}{\mathrm{v}}\}$, respectively. For the first term, we have

$$|(\nabla_b \overset{n}{\mathbf{T}}{}^{ab})\{\mathrm{w}_a\}| = |\overset{n}{\mathbf{J}}{}^b\{F_{ab}\mathrm{w}^a\}| = |\overset{n}{\mathbf{J}}{}^b\{\mathrm{s}_b\}| \leq \overset{n}{\mathbf{T}}\{\overset{n}{\mathrm{x}} + \overset{n}{\mathrm{y}}\}. \tag{15}$$

The first step uses the force law. For the second step, choose (as we always may) w_a to be tangent to γ on γ, whence $F_{ab}\mathrm{w}^a$ is orthogonal to u^a on γ. But every such test field can be written as a gradient (which is annihilated by $\overset{n}{\mathbf{J}}$, by conservation) plus a test field, s_b, that vanishes on γ. The third step, for all sufficiently large n follows from the fact that s_b vanishes on γ, and that the charge-mass ratio of $(\overset{n}{T}{}^{ab}, \overset{n}{J}{}^a)$ is bounded. The second term is converted, in a similar manner, to $\overset{n}{\mathbf{J}}\{\overset{n}{\mathrm{s}}\}$, where $\overset{n}{\mathrm{s}}_b$, again, is a test field vanishing on γ. But, in the proof of Theorem 3, $\overset{n}{\mathrm{v}}_b$, and so this $\overset{n}{\mathrm{s}}_b$, is chosen before we must choose $\overset{n}{\mathrm{y}}_{ab}$. So, using boundedness of the charge-mass ratio, we simply adjust our choice of $\overset{n}{\mathrm{y}}$, for each successive n, so that $\overset{n}{\mathbf{J}}\{\overset{n}{\mathrm{s}}\}$

of internal fluid states; so the mass and charge densities, μ and ρ, as well as the pressure p, are all functions on that manifold. We demand that there exist an open subset U of this 2-manifold such that i) the ratio ρ/μ is bounded in U; and ii) on the boundary of U, the gradient of p is tangent to that boundary. It follows from these conditions that, given any sample of this fluid with its internal state initially lying in U, then this sample will so evolve to always remain within U, i.e., will maintain in the course of its evolution a bounded ratio ρ/μ.

is bounded by, say, $(1/10)\overset{n}{\mathbf{T}}\{\overset{n}{\mathrm{X}} + \overset{n}{\mathrm{Y}}\}$. Incorporating these two bounds into (12), and making suitable adjustments in the numerical factors, the proof of Theorem 3 goes through as before. We conclude: There exists a sequence, $(\overset{n}{T}{}^{ab}, \overset{n}{J}{}^{a})$, each a multiple of an element of $\mathcal{C}$, such that the $\overset{n}{T}{}^{ab}$ converge to the distribution $u^a u^b \boldsymbol{\delta}_\gamma$.

Next, choose test vector field z_a satisfying $u^a \boldsymbol{\delta}_\gamma\{\mathrm{z}_a\} = 1$. It follows, from the bound on the charge-mass ratio and the fact that $\overset{n}{T}{}^{ab} \to u^a u^b \boldsymbol{\delta}_\gamma$, that the numbers $\overset{n}{\mathbf{J}}{}^{a}\{\mathrm{z}_a\}$ lie in a compact set. Hence, we may, taking a subsequence if necessary, assume that the $\overset{n}{\mathbf{J}}{}^{a}\{\mathrm{z}_a\}$ converge to some number, κ'. But every test field is equal to the sum of a multiple of z_a and a test field, s_a, that satisfies $u^a \boldsymbol{\delta}_\gamma\{\mathrm{s}_a\} = 0$, and, therefore, $\overset{n}{\mathbf{J}}{}^{a}\{\mathrm{s}_a\} \to 0$. It follows that $\overset{n}{J}{}^{a} \to \kappa' u^a \boldsymbol{\delta}_\gamma$.

Thus, under the requirement of bounded charge-mass ratio, the family $\mathcal{C}$ includes, up to a factor, bodies whose stress-energies approach that of a point mass while, furthermore, their charge-currents approach that of a point charge. But the $(\overset{n}{T}{}^{ab}, \overset{n}{J}{}^{a})$ satisfy the force law, and so therefore, taking the limit, must $(u^a u^b \boldsymbol{\delta}_\gamma, \kappa' u^a \boldsymbol{\delta}_\gamma)$. We conclude, then, that γ must be a Lorentz-force curve, with charge-mass ratio, κ', satisfying $|\kappa'| \leq \kappa$.

In this sense, then, charged bodies move on Lorentz-force curves. Again, we emphasize that we do *not* require that the (T^{ab}, J^a) converge to distributions on γ — and certainly not that they converge to any specific distributions. Rather, we only demand that the (T^{ab}, J^a) "collapse down" onto γ in the sense of tracking; and that, while doing so, they maintain bounded charge-mass ratio. It then *follows* that these fields converge to the distributions representing a point mass and point charge; and, further, that the curve γ have acceleration appropriate to such a particle.

We remark that the condition, in Theorem 4, that the members of $\mathcal{C}$ have a uniform bound on the charge-mass ratio, can be weakened. Indeed, all that is actually required in the proof of 4 is that the T^{ab} bound the J^a "on average". This could be expressed, not as pointwise inequalities on these fields, but rather as inequalities involving the results of applying them to certain test fields.

Is there a generalization of Theorem 3 to bodies subject to other forces, more general than electromagnetic? Consider a collection $\mathcal{C}$ of fields T^{ab}, each subject only to the energy condition. Each of these fields describes a body, where that body is subject to an effective force density, given by $\nabla_b T^{ab}$. Again, we shall need to exercise some control over this force. An obvious condition is that analogous to Eqn. (13) for the charge-current case: Demand that, for some positive number ϵ,

$$|(\nabla_b T^{ab}) t_a| \leq \epsilon T^{ab} t_a t_b, \tag{16}$$

for every unit timelike t^a at every point[9]. Here, $1/\epsilon$ represents, in physical terms, a lower limit on the time-scale over which the external forces can have a significant effect on the body.

It turns out, however, that this condition alone is not sufficient to achieve the conclusion of Theorem 3, for the following reason. In order to recover Eqn. (12), terms involving the force ($\nabla_b T^{ab}$, applied to certain test vector fields) must be bounded by terms involving the amount of matter present (T^{ab}, applied to certain test tensor fields). In the first step of Eqn. (12) for

[9] This condition is also easily generalized to distributions; and, so generalized, it implies that, for T^{ab} satisfying the energy condition, $\nabla_b T^{ab}$ must be order zero.

example, $(\nabla_b \overset{n}{\mathbf{T}}^{ab})\{\mathrm{w}_a\}$ must be bounded by $\overset{n}{\mathbf{T}}\{\overset{n}{\mathrm{X}} + \overset{n}{\mathrm{Y}}\}$. But here the test field w_a arises *after* we have made our choices of $\overset{n}{\mathrm{X}}$ and $\overset{n}{\mathrm{Y}}$. We can always choose w_a to be a multiple of u_a on γ, but, even with this further property, no bound of the type described in the previous paragraph will suffice.

Fix a unit timelike vector field, u^a, that, on γ, is the tangent to this curve. There is a simple physical reason why merely bounding $u_a \nabla_b T^{ab}$ by $T^{ab} u_a u_b$ does not suffice for Theorem 3. We may interpret $u_a \nabla_b T^{ab}$ as the rate of mass-transfer (as measured by u^a) to the body. But, if we allow mass-transfer to our body, then we cannot expect that the T^{ab} must converge to $u^a u^b \boldsymbol{\delta}_\gamma$, for the latter represents a particle of *constant* mass. Indeed, we have already seen these effects, for distributions, in Sect 2.

Such mass-transfer can arise in a variety of contexts. For example, for a star passing through a dust cloud, the rest mass of the star will increase due the accretion of dust. There are also more subtle examples. The act of striking a tennis ball will increase the rest-mass of that ball, for the stress created by the strike will, at least in part, be converted into heat within the ball. Indeed, it is difficult to think of any scenario in which external forces act on a body without the possibility, at least in principle, of mass-transfer.

Note that an external electromagnetic field, acting on a body carrying charge-current, can also result in mass-transfer, by the same mechanism as for the tennis ball. Why, then, did this issue not arise in our earlier treatment of charged bodies? The reason is that in that case we demanded that the charge-mass ratio of the material remain bounded — a very strong requirement. The mass-transfer in the electromagnetic case is driven by interaction between the dipole and higher moments of the body and the external field. But the bound on the charge-mass ratio ensures that these moments (per unit mass) go to zero in the limit, and so too must the mass-transfer they generate. This special feature of the electromagnetic case is reflected in the mathematics as follows. For w_a a test vector field tangent to γ on γ, the effective mass transfer — the result of applying the force density, $F^{ab} J_b$, to that test field — becomes, by virtue of conservation, J^b applied to a test field that *vanishes* on γ.

There are two possible lines to generalizing Theorem 3 to the case of more general forces.

For the first, we could strengthen the hypothesis of Theorem 3: We could simply demand that the mass-transfer (relative to the amount of matter present) vanish in the limit. That is, we could demand that $(\nabla_b T^{ab})$, applied to any test vector field that is tangent to γ on γ, be bounded by T^{ab}, applied to some test tensor field that vanishes on γ. Note that this condition has a very different character from those we considered above. Whereas our earlier conditions were imposed on the type of matter of which the bodies are composed, this condition is imposed on the manner in which those bodies are constructed.

For the second, we could weaken the conclusion of Theorem 3: We could conclude, not that the sequence $\overset{1}{T}{}^{ab}, \overset{2}{T}{}^{ab}, \cdots$, converge to the distribution $u^a u^b \boldsymbol{\delta}_\gamma$, but rather that, for some sequence of positive functions f_n, the result of multiplying each $\overset{n}{T}$ by that f_n converge to this distribution. This line, in other words, allows mass-transfer, but adjusts for it by adjusting the $\overset{n}{T}$ before taking the limit[10].

[10]There are also lines intermediate between these two. We could fix the mass-transfer along γ, once and for all. Then, to reflect this choice, we impose suitable conditions on the $u_a \nabla_b \overset{n}{T}{}^{ab}$ as well as on the factors by

These two lines are perhaps not all that different. There is no "action at a distance" in relativity: If you wish that forces be exerted on a body, you must introduce some other type of matter residing in the immediate vicinity of that body. Neither the stress-energy of the matter comprising the body, nor that of the matter in the environment, will be conserved, although of course their sum will be. It is this failure of these two types of matter to be separately conserved that results in a "force density" on the body. This scenario requires that the matter that is "part of the body" be distinguished from the matter that is "part of the environment", and this distinction may not always be clear-cut. Indeed, the same issue of making this distinction arose in Sect 2. These two lines, then, merely correspond to different ways of making this distinction. Indeed, the whole notion of a body, acted upon by external forces but otherwise maintaining its integrity, is perhaps not as natural in relativity as it is, say, in Newtonian mechanics.

Theorem 3 can be generalized to the null case. To this end, let γ be a null curve in space-time (M, g_{ab}), and let $\mathcal{C}$ be a collection of fields T^{ab} that are conserved, satisfy the energy condition, and track γ. Let t^a be a timelike vector field, defined on γ. We say that a choice, l^a, of tangent vector to γ is *affine* (with respect to t^a) provided $(l^m \nabla_m l^a) t_a = 0$. It is easy to check that such a tangent vector always exists, and that it is unique up to multiplication of l^a by a constant. In the special case in which γ is a (null) geodesic, the affine tangent vectors (with respect to t^a) are the usual geodesic affine tangents, independent of t^a. So, fix some t^a, as well as an affine tangent vector l^a with respect to that t^a. This choice of l^a generates a corresponding parameterization of the curve γ; and, with respect to that parameterization, $\boldsymbol{\delta}_\gamma$, the delta distribution of γ, makes sense. We have, for example, $\nabla_a(l^a \boldsymbol{\delta}_\gamma) = 0$. Now the proof of Theorem 3 goes through just as before, with l^a replacing u^a everywhere in that proof. For example: For z_{ab} any symmetric test field along γ satisfying $(l\, l \boldsymbol{\delta}_\gamma)\{\mathrm{z}\} = 0$, there does indeed exist a test field v_a such that $\mathrm{z}_{ab} - \nabla_{(a}\mathrm{v}_{b)}$ vanishes on γ (choosing for v_a a function times t_a). We conclude that some sequence $\overset{n}{T}{}^{ab}$, multiples of elements of $\mathcal{C}$, converge to $l^a l^b \boldsymbol{\delta}_\gamma$.

Thus, if a collection of conserved T^{ab} satisfying the energy condition tracks a null curve γ, then γ must be a null geodesic. Note that, quite generally, a collection $\mathcal{C}$ of T^{ab} satisfying the energy condition that tracks every timelike geodesic must also track every null geodesic.

In the treatment above, we always begin with an exact solution of Einstein's equation in which we have identified some material body; and we always end up with a space-time in which there is specified some timelike curve γ. Theorem 3 represents just one strategy to get from this beginning to this end: Introduce a family of bodies that, beginning with the given exact solution, have stress-energies that approach zero, after which the sizes of the bodies also approach zero.

But there are other strategies. Fix a space-time, (M, g_{ab}), together with a timelike curve γ in this space-time. Let $\mathcal{C}$ be a collection of smooth metrics on M, each of whose Einstein tensors satisfies the energy condition. Let us now demand: Given any neighborhood U of γ, any compact neighborhood C of a point of γ, and any C^0-neighborhood of the metric g_{ab} in $C - U$, there exists a metric g'_{ab} in the collection $\mathcal{C}$ such that i) its Einstein tensor is nonzero somewhere in $C \cap U$ and vanishes in $C - U$; and ii) the metric g'_{ab}, restricted to $C - U$, lies within the given C^0-neighborhood. This would seem to be the minimal arrangement that could be construed as

which the $\overset{n}{T}{}^{ab}$ are multiplied.

representing a family of bodies that "follow, in the limit, a curve γ." The bodies themselves are represented by the metrics in the collection $\mathcal{C}$, and the sense of "following" is reflected by condition on the Einstein tensors of these metrics. Note that we impose no conditions whatever on the internal construction of those bodies (i.e., in U): Their stress-energies can be large and can vary rapidly from point to point, and those stress-energies can produce large distortions of the space-time metric. We do, however, demand that the external metrics of these bodies approach a "background" in the sense that, in the $C - U$ (i.e., away from γ), those metrics C^0-approach some fixed metric g_{ab}.

We would not expect to be able to conclude, under this arrangement, that γ must be a geodesic: These bodies could, for example, propel themselves by emitting gravitational radiation. Clearly, there is a great deal of room between the conditions above — arguably, the weakest possible — and the very strong conditions that underlie Theorem 3. This suggests the following program: Start with the conditions above (which, apparently, do not restrict the final curve γ at all), and then, in order to conclude that that curve have various properties, impose additional conditions on the collection $\mathcal{C}$.

Here is an example. Let us strengthen the conditions above by demanding that the metrics in $\mathcal{C}$ C^1-converge to g_{ab} everywhere — that is, replace the C^0 neighborhood of g_{ab} in $C - U$, by a C^1 neighborhood of g_{ab} in all of C. Clearly, this stronger condition imposes a restriction also on the internal structure of the bodies. Indeed, it amounts, essentially, to the requirement that there be a universal upper bound to the their mass densities[11]. Thus, in the example of the earth in orbit around the sun, this condition contemplates a sequence in which the earth is replaced successively by a smaller planet, then by a rock, then by a grain of sand, etc.

It turns out that Theorem 3 can be adapted to apply under the condition above. Each metric g'_{ab} in $\mathcal{C}$ gives rise to a stress-energy, T'^{ab}. This T'^{ab} is, of course, conserved with respect to g', but with respect to g it manifests an effective force. That force necessarily satisfies Eqn. (16), and furthermore, by C^1-convergence, the $g'_{ab} \in \mathcal{C}$ can be so chosen that it further satisfies this equation for arbitrarily small ϵ. Now choose a sequence ϵ_n approaching zero sufficiently quickly, and then, in Theorem 3, choose each $\overset{n}{T}$ to satisfy (16) for that ϵ_n. The bound Eqn. (16) suffices to control the additional terms $\nabla_a \overset{n}{T}{}^{ab}$ that now arise in the first and fourth steps in Eqn. (12).

We thus conclude, from Theorem 3, that there is a sequence, $\overset{n}{T}{}^{ab}$, each a multiple of an element of $\mathcal{C}$, that approaches the distribution $u^a u^b \boldsymbol{\delta}_\gamma$. But approximate conservation of the $\overset{n}{T}$ produces, in the limit, exact conservation of this distribution. It follows that the timelike curve γ must be a geodesic. This is essentially the result of [14].

There may be other results along these lines.

4. Wave Packets

In this section we consider a class of examples, which will serve to illustrate the ideas discussed in Sects 2 and 3. In general terms, we consider wave packets composed of solutions of some system of partial differential equations. We are interested here in a limit in which the

[11] A body of mass density ρ and size L distorts the metric by the order of ρL^2; and the derivative operator by the order of ρL. These go to zero as $L \to 0$, provided ρ bounded.

wave packet becomes both smaller and longer-lived. That is, we are interested in a limit in which the packet as a whole follows some curve in space-time.

Fix, once and for all, a globally hyperbolic space-time (M, g_{ab}). We impose global hyperbolicity here solely in order to guarantee that the solutions of our equations are sufficient in both number and diversity. It may be that some weaker condition on the space-time will suffice. Next, consider a system of linear partial differential equations on some fields on this space-time. We suppose that we are given an expression for a stress-energy tensor, T^{ab}, quadratic in those fields, and that this stress-energy, by virtue of its construction, automatically satisfies the energy condition. We do *not* demand that this T^{ab} be conserved: There may be external forces acting, through the equations, on these fields. Next, fix a timelike or null curve, γ, in this space-time. Then: Some given collection, $\mathcal{S}$, of solutions of this system of equations will be said to *track* γ provided the collection of stress-energies, T^{ab}, computed from those fields tracks γ in the sense of Sect 3.

An example of what we have in mind is the Maxwell system. Here, we have an antisymmetric tensor field, F_{ab}, subject to Maxwell's equations (say, with zero sources): $\nabla^a F_{ab} = 0$, $\nabla_{[a} F_{bc]} = 0$. The stress-energy of this field, given by $T^{ab} = F^a{}_m F^{bm} - 1/4 g^{ab} F^{mn} F_{mn}$, satisfies the energy condition and (by virtue of Maxwell's equations) is conserved. Let the collection $\mathcal{S}$ consist of *all* solutions of Maxwell's equations in this space-time. Which causal curves γ does this collection track? Since the stress-energy is conserved in this case, it follows from Sects 2 and 3 that the only candidates for such curves are the (timelike or null) geodesics. It turns out that this $\mathcal{S}$ tracks no timelike geodesics. To see this, apply a conformal rescaling to this space-time, i.e., replace the metric g_{ab} by $\Omega^2 g_{ab}$, where Ω is some smooth positive function on the manifold. Every such rescaling preserves Maxwell solutions — and therefore the curves that the collection $\mathcal{S}$ tracks — but these rescalings in general fail to preserve the geodesic character of timelike curves. We conclude, then, that it is only the null geodesics that remain as viable candidates for those our collection $\mathcal{S}$ tracks.

In fact, the collection $\mathcal{S}$ tracks *every* null geodesic in the space-time (M, g_{ab}). This is most easily seen for Minkowski space-time. Fix a null geodesic γ, and let x_{ab} be a test field that satisfies the dual energy condition in a neighborhood of γ and is generic at some point of γ. Fix a point p of γ, sufficiently far in the past along this curve that there is some neighborhood U of p that does not meet the future of the support of x_{ab}. Then a Maxwell field, generated by initial data supported in $U \cap I^+(p)$, will meet the support of x_{ab} only in $I^+(p)$, where $I^+(p)$ denotes the future of p. Furthermore, the center of mass of this field will be a timelike geodesic passing through U. Now consider a sequence of such fields, generated by initial data supported in successively smaller neighborhoods of p, and apply to these successively larger boosts that preserve both γ and p. There results a sequence of Maxwell solutions, each meeting the support of x_{ab} only in $I^+(p)$, such that their centers of mass converge to γ. Clearly, the stress-energies of the solutions in this sequence will eventually satisfy $\mathbf{T}^{ab}\{\mathrm{x}_{ab}\} > 0$. That is, this sequence, and so $\mathcal{C}$ itself, tracks γ.

This result is easily generalized to curved space-time. Let (M, g_{ab}) be a space-time with Cauchy surface S, and let γ be any null geodesic in this space-time. Fix a neighborhood U of γ, a function f on M having value 1 in some neighborhood of γ and vanishing outside of U, and a flat metric, g^0_{ab}, defined in U, such that g_{ab} and g^0_{ab}, together with their first derivatives, agree on γ. Finally, fix an isometric embedding of (U, g^0_{ab}) in Minkowski space-time, and denote by γ'

the image of γ under this embedding, so γ' is also a null geodesic. Now, given any solution F'_{ab} of Maxwell's equations in Minkowski space-time, set Set $F = F_1 + F_2$, where F_1 is the result of pulling F' back to U via the embedding and multiplying by f; and F_2 is the Maxwell field in (M, g_{ab}) that vanishes on S and has sources given by $(-\nabla_b F_1^{ab}, -\nabla^{[a} F_1^{bc]})$. Then this F_{ab} is a source-free solution of Maxwell's equation in (M, g_{ab}). Suppose, next, that the F' track γ' in the Minkowski space-time. Then, we claim, the corresponding F's track γ in (M, g_{ab}). Indeed, the F_1 clearly track γ. But we also have a bound on the sources for F_2, as follows from the fact the F' satisfy Maxwell's equations in the Minkowski space-time and track γ' there, together with the defining properties of f and g^0_{ab}. It follows from this bound that the contribution of F_2, relative to F_1, can be made as small as we wish.

We conclude, then, that the collection $\mathcal{S}$ of all solutions of Maxwell's equations in any globally hyperbolic space-time tracks every null geodesic γ in that space-time — these curves and no other curves. This conclusion reflects what is usually called the "optical limit" of electromagnetism. We remark that the present formulation of the optical limit is precise and remarkably simple. The key idea that makes this happen is the notion of tracking.

We turn next to a second example — the Klein-Gordon equation. Fix a positive number m. Then the Klein-Gordon field is a complex scalar field ϕ on M, subject to the equation $\nabla^2\phi - m^2\phi = 0$. The stress-energy of this field is given by

$$T^{ab} = \nabla^{(a}\phi\nabla^{b)}\overline{\phi} - (1/2)(\nabla^n\phi\nabla_n\overline{\phi})g^{ab} - (1/2)m^2\phi\overline{\phi}g^{ab}. \tag{17}$$

This T^{ab} satisfies the energy condition, and, again, is conserved. Again, we let $\mathcal{S}$ be the collection of *all* solutions of the Klein-Gordon equation (for this fixed value of m) in our space-time (M, g_{ab}), and again we ask for those curves γ that this collection tracks. It follows, again from conservation, that the only candidates are the timelike and null geodesics.

We first note that, in the case (M, g_{ab}) Minkowski space-time, the collection $\mathcal{S}$ does in fact track every null geodesic, by the same argument as for the Maxwell case. This is what we would have expected. Think of a Klein-Gordon wave packet as representing a massive particle. In the high-energy limit, such a particle would nearly follow a null geodesic; and so we expect that the corresponding wave packets would track those curves.

But massive particles at lower energies typically follow timelike geodesics. Does the collection $\mathcal{S}$ track these curves, too? It turns out that it does not. To see this, suppose, for contradiction, that $\mathcal{S}$ did track some timelike geodesic, γ. Then, by Theorem 3, some sequence of stress-energies, (17), must converge, as distributions, to $u^a u^b \boldsymbol{\delta}_\gamma$. But this in turn requires, from Eqn. (17), that $\nabla^{(a}\phi\nabla^{b)}\overline{\phi}$ and $\nabla^n\phi\nabla_n\overline{\phi} + m^2|\phi|^2$ converge to $u^a u^b \boldsymbol{\delta}_\gamma$ and zero, respectively. It follows that $|\phi|^2$ must converge to $\boldsymbol{\delta}_\gamma/m^2$; and, therefore, that $\nabla^2(|\phi|^2)$ must converge to $\nabla^2(\boldsymbol{\delta}_\gamma/m^2)$. But the former is equal to $2\nabla^n\phi\nabla_n\overline{\phi} + 2m^2|\phi|^2$, which, as we have just seen, converges to zero. We now have a contradiction, for $0 \neq \nabla^2(\boldsymbol{\delta}_\gamma/m^2)$. In short, the Klein-Gordon stress-energy (17) has the wrong "shape" to give rise, in the limit, to a point-particle mass distribution[12].

This conclusion is what we would expect geometrically. Think of a Klein-Gordon wave packet in Minkowski space-time as composed of plane waves, each with a frequency-wave number vector, k^a, satisfying $k^a k_a = -m^2$. In order that such a packet be long-lived, it must be

[12]We remark that a similar argument gives an alternative proof that the solutions of Maxwell's equations track no timelike curve.

the case that the bulk of the waves comprising that packet have k-values close to some fixed vector, k_0^a. And, in order that the packet itself be small in size, it must be the case that the wavelength associated with this k_0 be small. But a frequency-wave number vector k_0, of fixed norm, can reflect small wavelengths only if it lies near the light cone. In physical terms, the parameter m that appears in the Klein-Gordon equation is related to the physical mass by a factor of Planck's constant, $\hbar$. It is only in the classical limit, $\hbar \to 0$, that we expect wave packets to track curves in the spacetime. But this limit corresponds, for fixed physical mass, to $m \to \infty$.

It is clear, from the discussion above, that Klein-Gordon solutions become more and more efficient at forming wave packets as the Klein-Gordon mass, m, of those solutions increases. This observation suggests that we proceed as follows. Let the collection $\mathcal{S}$ consist, not of the Klein-Gordon solutions for some fixed value of the parameter m, but rather of all solutions for *all* values of this parameter.

This collection $\mathcal{S}$, it turns out, does indeed track every timelike geodesic. To see this, consider first the case in which (M, g_{ab}) is Minkowski space-time. Fix a "time function" t in this space-time, i.e., such that $u^a = \nabla^a t$ is constant, unit, and timelike. For each value of t, denote by S_t the spacelike 3-surface of constant t. For ζ any smooth, complex-valued function on M, denote by $E(\zeta, t)$ the value of the integral of $T_\zeta{}^{ab} u_b$ over the surface S_t, where $T_\zeta{}^{ab}$ denotes the result of replacing ϕ by ζ in Eqn. (17). Then $E(\zeta, t) \geq 0$; and we have, by direct computation, $\nabla_b T_\zeta{}^{ab} = 1/2[\nabla^a \zeta(\nabla^2 - m^2)\overline{\zeta} + \nabla^a \overline{\zeta}(\nabla^2 - m^2)\zeta]$. It follows that

$$dE(\zeta, t)/dt \leq 2E(\zeta, t)^{1/2}\left(\int_{S_t} |(\nabla^2 - m^2)\zeta|^2 dS\right)^{1/2}. \tag{18}$$

Think of $E(\zeta, t)$ as an "effective energy" of the function ζ, and of Eqn. (18) as reflecting the idea that this energy can grow only to the extent that the function ζ fails to satisfy the Klein-Gordon equation.

Next, fix any static (with respect to u^a), complex-valued function α, of compact spatial support, on the manifold M. Denote by ϕ the solution of the Klein-Gordon equation, for some value of $m > 0$, whose initial data, on the surface S_0, are $(\alpha|_{S_0}, im\alpha|_{S_0})$; and by ϕ' the function $\alpha \exp(imt)$ on M. The two functions ϕ and ϕ' manifest the same initial conditions on S_0, and furthermore their difference satisfies $(\nabla^2 - m^2)(\phi - \phi') = \exp(imt)\nabla^2\alpha$. Setting $\zeta = \phi - \phi'$ in Eqn. (18), there follows a bound on $E(\phi - \phi', t)$, independent of m. In the region of S_t outside of the support of α, ϕ' vanishes; and so this bound also serves as a bound, still independent of m, for the energy-integral of ϕ, taken only over this region. But the total energy of ϕ is given by $E(\phi, t) = (1/2)\int_{S_0}(m^2 \alpha\overline{\alpha} + \nabla^a \alpha \nabla_a \overline{\alpha})dA$, which is independent of t, and grows without bound as $m \to \infty$. We conclude: The value of the energy integral of ϕ, taken over the entire surface S_t dominates the value of that energy taken only outside the support of α — and the extent of this domination increases as m increases. It follows that the collection of Klein-Gordon solutions ϕ, constructed as above (for all $m > 0$), track every timelike geodesic orthogonal to the S_t. Similarly for other timelike geodesics in Minkowski space-time.

This result is easily generalized to curved space-times, by an argument similar to that for the Maxwell case. We conclude, then, that, in any globaly hyperbolic space-time, the collection $\mathcal{S}$ of all Klein-Gordon solutions for all values of $m > 0$ tracks precisely the timelike and null geodesics in that space-time.

We next turn to the case of the charged Klein-Gordon field. Fix, on the space-time (M, g_{ab}), an antisymmetric tensor field F_{ab}, the background field generated by some external sources. Fix also numbers m and e. Then a Klein-Gordon field (for these values of m and e) is a charge-e scalar field ϕ (necessarily complex), subject to the equation $\nabla^2\phi - m^2\phi = 0$, where ∇_a is the charge-derivative operator. The stress-energy of this field (which, again, satisfies the energy condition) is given by (17), and the charge-current by

$$J^a = (ie/2)(\overline{\phi}\nabla^a\phi - \phi\nabla^a\overline{\phi}). \tag{19}$$

There follows: $\nabla_b J^b = 0$ (charge conservation) and $\nabla_b T^{ab} = F^a{}_b J^b$ (force equation). Thus, the stress-energy in this case fails to be conserved: There is an external force, the Lorentz force, acting on the charged fields. Note that, for ϕ a charged Klein-Gordon solution, for some values (e, m), then $\overline{\phi}$ is also a solution, for $(-e, m)$; and these two solutions have the same stress-energy and charge-current. It is immediate from (19) and (17) that the charge-mass ratio for a charged Klein-Gordon field is bounded, in the sense of Sect 3, by the number $|e|/m$.

Fix, once and for all, a globally hyperbolic space-time, (M, g_{ab}), and a number κ. Denote by $\mathcal{S}_\kappa$ the collection of all Klein-Gordon solutions on this space-time with $m > 0$ and $e/m = \kappa$. This collection includes solutions for arbitrarily large values of m: As we have seen in the uncharged case, it is only in the limit $m \to \infty$ that interesting tracking behavior emerges. Let γ be a timelike curve in this space-time.

Suppose, in the first instance, that the collection $\mathcal{S}_\kappa$ tracks γ. It then follows, from Theorem 4, that γ must be a Lorentz-force curve, with charge-mass ratio in the closed interval $[-\kappa, \kappa]$.

Next, let there be given a curve γ, a Lorentz-force curve with charge-mass ratio κ. Then, we claim, the collection $\mathcal{S}_\kappa$ tracks γ. This follows by essentially the same argument as in the uncharged case. (Use a choice of vector potential for the Maxwell field F_{ab}, to convert charged scalar fields to ordinary (complex) scalar fields; and the charged derivative operator to the ordinary derivative operator, corrected by a term involving the vector potential.) If, in that argument, we replace im by $-im$, we obtain wave packets that track the Lorentz-force curves with charge-mass ratio $-\kappa$. (Indeed, it is easy to prove, quite generally, that S_κ and $S_{-\kappa}$ track precisely the same curves, as a consequence of the fact that the operation of complex conjugation, which sends S_κ to $S_{-\kappa}$, preserves tracking.)

To summarize, we have shown that every timelike curve tracked by the collection $\mathcal{S}_\kappa$ of charged Klein-Gordon fields is a Lorentz-force curve with charge-mass ratio in the interval $[-\kappa, \kappa]$; and, conversely, that every curve whose charge-mass ratio lies at either endpoint of this interval is, indeed, tracked by $\mathcal{S}_\kappa$. Are the Lorentz-force curves with charge-mass ratio in the interior of this interval also tracked? We suspect that they are not, for the following reason. Fix, say, in Minkowski space-time, an initial surface S_0. Then initial data on S_0 of the form $(\alpha, im\alpha)$, where α is any function on S_0, give rise, as $m \to \infty$, to a packet that follows a Lorentz-force curve with charge-mass ratio κ; while initial data of the form $(\alpha, -im\alpha)$ gives rise to a packet that follows a curve with ratio $-\kappa$. But *every* set of initial data on S_0 can be written, uniquely, as the sum of one set data of the first form and one of the second.

The general conclusion, in any case, is that charged Klein-Gordon fields, in an appropriate limit, follow the corresponding Lorentz-force curves. It is curious that quantum field theory exploits an entirely different mechanism to achieve Lorentz-force motion. Consider charged, spin-zero particles in Minkowski space-time. The Hilbert space of one-particle states is formed, not from the charged-field solutions of the Klein-Gordon equation, but rather from the charge-zero solutions. The effect of an external electromagnetic field is represented by an interaction on this Hilbert space, and it is this interaction that, in the classical limit, is responsible for Lorentz-force motion.

Section 5. Conclusion

The theory of a "point particle" — represented by a distribution $\mathbf{T}^{ab}$ supported on a timelike curve γ — is remarkably simple. Here is a context in which the motion of a body makes sense. And, indeed, we recover, in this context, what we expect: that γ is a geodesic (in the case of no external force), or a Lorentz-force curve (for a particle carrying charge). We have argued that this model is a reliable indicator of how actual, extended bodies will behave: If we demand of a body only that it be "sufficiently small", in both size and mass, then that body is already well-represented by the corresponding distribution. The key notion here is that of tracking. Indeed, tracking applies even to "bodies" constructed as wave packets of, e.g., Maxwell or Klein-Gordon fields. This leads, among other things, to a simple, transparent version of the optical limit for electromagnetism.

There remain a number of open issues. What is the mechanism by which a particle achieves the torque-condition, (5) (or, in the electromagnetic case, (10))? How, for example, does this condition emerge from the dynamics of extended bodies? For a particle subject to an external force, one can exploit the freedom to exchange matter between the particle and its environment to simplify the stress-energy, leading to Eqns. (6)-(7). Is there a similar freedom for an extended body; and how does it operate in the limit? Also, we gave a physical argument that these equations should require that the relevant distributions — $\boldsymbol{\mu}$ and $\mathbf{f}^a$ — be multiples of $\boldsymbol{\delta}_\gamma$. Can this argument be placed on firmer footing? Finally, we noted that Eqns. (6)-(7), while they do not have a meaningful initial-value formulation in general, do have such a formulation in the electromagnetic case. What is the status of this issue for other forces? We saw that a particle carrying charge cannot manifest electromagnetic multipole moments higher than dipole, resulting in Eqn. (8). Could one see in more detail how this happens, by considering a limit of extended bodies? Can Theorem 3 be generalized to forces other than electromagnetic (e.g., contact forces)? Can it be generalized to include other strategies by which particles emerge as limits of extended bodies?

Here is a curious example of a further application of the notion of tracking. We would certainly expect that general relativity will, in some sense, prohibit "tachyonic bodies", i.e., those that follow spacelike curves. Tracking, it turns out, provides a precise formulation of this idea. We claim: In any space-time, the collection $\mathcal{C}$ of all conserved T^{ab} satisfying the dominant energy condition tracks no spacelike curve. To see this, suppose, for contradiction, that $\mathcal{C}$ tracked some spacelike curve γ. Let v_a be a test vector field such that $\nabla_{(a}\mathrm{v}_{b)}$ satisfies the dual energy condition in a neighborhood of γ and is generic at some point of γ. [To construct such a field, fix, in a neighborhood of γ, a time function t that is positive on all but a finite segment of γ, and set, in that neighborhood, $\mathrm{v}_a = f(t)\nabla_a t$, where f is a suitable smooth, nonnegative function of one variable that vanishes for $t \geq 0$.] We now have a contradiction, for

$\mathbf{T}^{ab}\{\nabla_{(a}\mathrm{v}_{b)}\}$ must vanish for all T^{ab} in $\mathcal{C}$, by conservation; but must be positive for some such T^{ab}, by tracking. It turns out that the dominant energy condition is essential for this argument. Indeed, in Minkowski space-time, for example, the collection of conserved T^{ab} satisfying the weak energy condition tracks (with that definition suitably adapted to that energy condition) every spacelike geodesic. This follows, noting that every $T^{ab} = \mu x^a x^b$, where x^a is a constant spacelike vector field and the function μ is nonnegative and constant along the x-trajectories, is conserved and satisfies the weak energy condition.

It is of some interest to ask which of the present results hold also for Newtonian gravitation. Recall that, in that theory, there is specified a 4-manifold of events, on which there is given (among other things) a function t (the "Newtonian time") and a derivative operator ∇_a (which includes the effects of gravitation). Matter, as in relativity, is described by a conserved, symmetric stress-mass-momentum tensor, T^{ab}. The natural candidate for an energy condition on such a tensor is the requirement that the combination $T^{ab}\nabla_a t \nabla_b t$ (the "mass density") be nonnegative. This condition differs in an important way from the dominant energy condition in relativity: Whereas the latter (which involves T^{ab} contracted with a variety of vectors) controls the entire stress-energy tensor, the former (since T^{ab} is contracted only with the single vector $\nabla_a t$) leaves certain components of T^{ab} unrestricted.

Consider, in a Newtonian space-time, a curve γ, parameterized by t, and a conserved distributional $\mathbf{T}^{ab}$, satisfying this energy condition and supported on γ. Then it is easy to prove that $\mathbf{T}^{ab}\nabla_a t \nabla_b t$ must be a multiple of $\boldsymbol{\delta}_\gamma$; and that, if this multiple is nonzero, then γ must be a geodesic. But (in contrast to the case in relativity) the remaining components of the distribution $\mathbf{T}^{ab}$ remain essentially free, and, indeed, can be of arbitrarily high order. There is no natural generalization of the full Theorem 3 to Newtonian gravitation, because the energy condition in that theory permits mass to be transported, arbitrarily rapidly, from place to place. There is, however, one result [15] along the lines of Theorem 3: Given any curve γ tracked by a collection $\mathcal{C}$ of fields T^{ab} (under a slightly stronger energy condition than that above, and under a further condition on the supports of the T^{ab}), then that curve must be a geodesic.

In Sect 4, we considered wave packets constructed of Maxwell and Klein-Gordon fields, and determined which curves these fields track. To what extent can these results be extended to more general systems of equations?

Consider, for example, Yang-Mills fields. In a fixed, globally hyperbolic background space-time, let $\mathcal{S}$ denote the collection of Yang-Mills fields with zero source. Then tracking makes sense (since the Yang-Mills stress-energy satisfies the energy condition); and any timelike or null curve this $\mathcal{S}$ tracks must be a geodesic (since this stress-energy is conserved). We would expect that it is precisely the null geodesics that are tracked, although the proof will be considerably more difficult than in the Maxwell case, because the Yang-Mills equation is nonlinear. Next, fix a background Yang-Mills field in this space-time, and consider mass-m boson fields carrying Yang-Mills charge. The stress-energy of such a field satisfies the energy condition, as well as a force-equation involving the Yang-Mills current of that field. We would expect behavior similar to that of Klein-Gordon: That the collection of such fields, for fixed m, tracks the null geodesics; and that that collection, for all $m > 0$, tracks also those timelike curves manifesting Yang-Mills force.

Consider, as a second example, the Dirac system. Recall that the Dirac field (of charge e, mass m) is given by a pair, $(\xi^A, \eta^{A'})$, of charge-e spinor fields, satisfying

$$\nabla_{AA'}\xi^A = (m/\sqrt{2})\eta_{A'}, \quad \nabla_{AA'}\eta^{A'} = (m/\sqrt{2})\xi_A. \tag{20}$$

The stress-energy and charge-current of this Dirac field are given by

$$T^{ab} = (i/2)(\xi^A \nabla^b \overline{\xi}^{A'} - \overline{\xi}^{A'} \nabla^b \xi^A + \eta^{A'} \nabla^b \overline{\eta}^A - \overline{\eta}^A \nabla^b \eta^{A'} + (a \leftrightarrow b)), \tag{21}$$

$$J^a = e(\xi^A \overline{\xi}^{A'} + \eta^{A'} \overline{\eta}^A). \tag{22}$$

where "$a \leftrightarrow b$" means the same expression, but with the roles of "a" and "b" reversed.

Note that for the Dirac field, in contrast to the Klein-Gordon field, the stress-energy does not in general satisfy the energy condition[13]. Consequently, we do not have the notion of a family of Dirac solutions tracking a curve, in the sense of Sect 3.

In any case, we might expect that there can be constructed, from solutions of the Dirac equation, wave packets (suitably defined) that follow timelike curves. Fix a number κ. Then, we might expect that, from the collection of all solutions of the Dirac equation with charge-mass ratio κ, there will be a sequence that follows (in some suitable sense) any Lorentz-force curve with charge-mass ratio κ. Why merely a Lorentz-force curve? Why do we not expect, in the equation for this curve, an additional term involving the spin of the particle interacting with the Riemann tensor of the space-time (M, g_{ab})? The reason is the following. The parameters m and e that appear in the Dirac equation are in geometrized units — related to the physical mass and charge by factors involving Planck's constant $\hbar$. In order to generate wave packets that follow curves, we must have $e, m \to \infty$, and this, for fixed physical charge and mass, requires $\hbar \to 0$. Thus, we are contemplating a limit in which the spin of the particle (relative to its mass), and so the effect of that spin on the particle's motion, goes to zero.

Strangely enough, there is a different notion of tracking available in the Dirac case. Let, say, $e > 0$, so the charge-current J^a is future-directed causal. Let us say that a collection of Dirac solutions tracks, in this new sense, curve γ provided: Given any test field x_a, past-directed causal in a neighborhood of γ and timelike at some point of γ, there is a field in this collection with $\mathbf{J}^a\{\mathrm{x}\} > 0$. It turns out that the proof of Theorem 3 goes through (in fact, somewhat more easily) with this new notion of tracking: It follows that there is some sequence from $\mathcal{C}$ with the J^a converging, up to rescaling, to $u^a \boldsymbol{\delta}_\gamma$. Unfortunately, we cannot conclude from this that γ must be a geodesic.

Are there similar results regarding the tracking behavior of solutions of general symmetric-hyperbolic systems of equations?

Appendix A: Distributions

We review briefly a few facts about distributions.

[13]The easiest way to see this is to note that there is a complex-conjugation operation on Dirac fields, similar to that on Klein-Gordon fields, but in this case it reverses the signs of both the stress-energy and the charge-current. Restricting consideration to positive-frequency solutions of the Dirac equation in Minkowski space-time will not work, for even this restricted class of solutions fails to satisfy the energy condition.

Fix, once and for all, a smooth manifold M. By a *test field* on M, we mean a smooth tensor density (of weight 1 – See Appendix B) of compact support. The test fields, of given index structure, say $\mathrm{T}^b{}_{ac}$, form a vector space. A (tensor) *distribution*, $\mathbf{d}_b{}^{ac}$, is a linear map, from this vector space to the reals, that is continuous in the following sense: Given any $\epsilon > 0$ and compact subset C of M, there exists an integer $n \geq 0$ and a number $\delta > 0$ such that $|\mathbf{d}\{\mathrm{T}\}| \leq \epsilon$ for every test field T whose support lies in C and which, together with its first n derivatives, is everywhere less than or equal to δ. In this definition, the "sizes" of the test field and its derivatives are measured using any fixed positive-definite metric and any derivative operator on M (the resulting notion of "continuous" being, of course, independent of these choices). Here, and hereafter, we use "{ }" to denote the action of a distribution on a test field. Distributions with other index structures are defined similarly. Indeed, distributions can be defined for any kind of fields on M, provided only that those fields have vector-space structure at each point. For example, there are distributions based on spinor fields, on charged fields, on densities, etc. In the case in which there is given a fixed metric, g_{ab}, on the manifold M, then we may represent test fields on M by ordinary tensor fields, converting these to densities using the alternating tensor of g_{ab}.

Note that every smooth tensor field gives rise to a distribution, where the action of the map is given by contracting that field with the test field and integrating over M. In fact, every merely continuous (or even somewhat less well-behaved) tensor field gives rise to a distribution.

We can add two distributions (having the same index structure), contract a distribution, and take the outer product of a distribution with a smooth tensor field. For example, the outer product of distribution $\mathbf{d}_b{}^{ac}$ and smooth vector field w^d is that distribution which, applied to test field $\mathrm{T}_d{}^b{}_{ac}$, yields the number $\mathbf{d}\{w^d \mathrm{T}_d{}^b{}_{ac}\}$. These operations are indicated in the usual way (repeated indices for contraction, juxtaposition for outer product). Furthermore, given any derivative operator, ∇_a, on M, that operator can be extended to act on distributions: For $\mathbf{d}_b{}^{ac}$ a distribution, $\nabla_m \mathbf{d}_b^{ac}$ is defined as that distribution which, applied to test field $\mathrm{T}^{mb}{}_{ac}$, yields the number $\mathbf{d}_b{}^{ac}\{-\nabla_m \mathrm{T}^{mb}{}_{ac}\}$. (This is the formula suggested by "integration by parts".) Note that *every* distribution, no matter how badly behaved, is (infinitely) "differentiable". Applied to smooth tensor fields, regarded as distributions, these operations reduce to the usual algebraic and differential operations on those tensor fields. And, applied to distributions quite generally, these operations satisfy all the usual properties. So, for example, given any system of linear partial differential equation on some tensor fields, there is a corresponding system of linear partial differential equation on the corresponding distributions. Thus, we have a version of Maxwell's equations with distributional Maxwell field and charge-current.

The *support* of a distribution $\mathbf{d}$ is the smallest closed set $C \subset M$ such that $\mathbf{d}\{\mathrm{T}\} = 0$ for every test field T whose support does not intersect C. So, for example: The support of the derivative of a distribution is a subset of the support of that distribution. We say that a distribution is of *order* n (where n is a nonnegative integer) provided the action of $\mathbf{d}$ can be extended, from (C^∞) test fields (of compact support), to C^n fields of compact support, continuous in the obvious topology. Thus, every distribution of order n is automatically of every order $\geq n$; and the derivative of such a distribution is automaticaly of order $n+1$. Note that the outer product of an order-n distribution and a C^n tensor field makes sense, and is itself an order-n distribution. Every distribution of compact support (but not necessarily those of non-compact support) has some (finite) order. Note that a zero-order (but not in general a higher-order) distribution annihilates every test field that vanishes on its support. Every distribution arising from a smooth tensor field has order zero, and support given by the support of that tensor field.

As an example, let (M, g_{ab}) be a space-time, and γ a smooth timelike curve therein. Denote by $\boldsymbol{\delta}_\gamma$ the scalar distribution that assigns, to test field T, the result of first converting T to a scalar field (using the alternating tensor of g_{ab}) and then integrating (with respect to length) over γ. This $\boldsymbol{\delta}_\gamma$ is called the *delta-distribution* of γ. It has order zero and support given by the curve itself. It further satisfies $\nabla_a(\boldsymbol{\delta}_\gamma u^a) = 0$, where ∇_a is the derivative operator determined by g_{ab}, and u^a is any smooth vector field that, on γ, is the unit tangent to that curve. Indeed, $\boldsymbol{\delta}_\gamma$ is, up to a constant factor, the unique order-zero distribution, with support on γ, with this property. The following is a useful fact: For α any smooth tensor field on M, the distribution $\nabla_a(\alpha u^a \boldsymbol{\delta}_\gamma)$ depends only on the values of α at the points of γ, while the distributions $\alpha u^a \nabla_a \boldsymbol{\delta}_\gamma$ and $u^a \nabla_a(\alpha \boldsymbol{\delta}_\gamma)$ do not in general have this property. Note that there is no analogous "delta distribution" for a general null curve.

There is a natural topology on the vector space of distributions of fixed index structure. To specify a neighborhood of a distribution $\mathbf{d}$, fix a finite list, $\mathrm{T}_1, \mathrm{T}_2, \cdots, \mathrm{T}_n$, of test fields, and a number $\epsilon > 0$. Then the neighborhood consists of all distributions $\mathbf{d}'$ such that $|\mathbf{d}\{\mathrm{T}_i\} - \mathbf{d}'\{\mathrm{T}_i\}| \leq \epsilon$ for all $i = 1, 2, \cdots, n$. Thus, for example, if p_i is a sequence of points of M, converging to $p \in M$, then $\boldsymbol{\delta}_{p_i}$, the sequence of delta-distributions of those points, converges in this topology to $\boldsymbol{\delta}_p$. Furthermore: If a sequence of distributions $\mathbf{d}_1, \mathbf{d}_2, \cdots$ converges to some distribution $\mathbf{d}$, then the sequence $\nabla_a \mathbf{d}_1, \nabla_a \mathbf{d}_2, \cdots$ converges to $\nabla_a \mathbf{d}$. We remark that every distribution on M is a limit, in this topology, of a sequence of distributions that arise from smooth tensor fields on M.

Appendix B: Tensor Densities

The test fields on which distributions act are densities. In the present context, we wish to impose on these test fields certain inequalities, such as a version of an energy condition. But for densities as normally defined, inequalities of this sort make no sense, because of the ambiguity as to the sign of the alternating tensor. For present purposes, therefore, it is convenient to introduce a slightly different notion of a density.

Fix a smooth (say, 4-dimensional) manifold M, not necessarily orientable. Fix also a point p of M. A *tensor density* (for order 1) at p is a pair, (t, ϵ), where t is a tensor at p and ϵ is a nonzero, fourth-rank, covariant, totally antisymmetric tensor at p; and where we identify each such pair with all other pairs of the form $(|a|t, a^{-1}\epsilon)$, as a runs over the nonzero reals. A (smooth) density field is a smooth assignment of a density (of given index structure for t) to each point of M.

Note that a density, as here defined, is slightly different from what is usually called a "density". For example, it makes sense to say that a scalar density, under the present definition, is "≥ 0" (for the demand that $t \geq 0$ is preserved under $(t, \epsilon) \to (|a|t, a^{-1}\epsilon)$). Note that every manifold M, even a non-orientable one, admits some positive scalar density field.

We can add two density fields having the same index structure (by choosing representatives at each point having the same underlying ϵ, and then adding the tensors t of those representatives). We can also contract a density field, and take the outer product of a density field and a tensor field, in the obvious way. These algebraic operations have all the standard properties. Thus, for example, the density fields, of given index structure, form a vector space. Finally, given any derivative operator, ∇_a, on tensor fields on M, we can extend that operator to act also on density fields, the result of this operation again being a density field. And this operator on density fields has all the usual properties, such as linearity and the Leibnitz rule.

Let α be a smooth scalar density field on M, and let $O \subset M$ be open. Then $\int_O \alpha$ makes sense (provided the integral converges). Indeed, if O is orientable, choose any alternating tensor field on O, and use it to convert α into an ordinary scalar field and also as the volume element to integrate that scalar over O. The result is independent of the choice of alternating tensor. If O is non-orientable, write it as a union of orientable regions, and define the integral over O as the sum of the integrals over these regions. Note that this integral, so defined, does *not* require that the manifold M be orientable, nor that there be specified any volume-element on M. As an example, we have: If $\alpha \leq 0$, then $\int_O \alpha \leq 0$.

Finally, consider the case in which there is specified some fixed Lorentz metric g_{ab} on M (which, still, may be non-orientable). Then this g_{ab} gives rise to an alternating tensor, ϵ_{abcd}, at each point of M, up to sign. We may now use this $\pm\epsilon_{abcd}$ to convert each tensor density field on M into an ordinary tensor field (of the same index structure). Note that a nonnegative scalar density is thereby converted to a nonnegative scalar field.

References

[1] A. Einstein and J. Grommer, *Allgemeine Relativitätstheorie und Bewegungsgesetz*. Berlin: Verlag der Akademie der Wissenschaften, 1927.

[2] A. Einstein, L. Infeld, and B. Hoffman, "The gravitational equations and the problem of motion," *Annals of Mathematics*, vol. 39, no. 1, pp. 65–100, 1938.

[3] H. Asada, T. Futamase, and P. A. Hogan, *Equations of motion in general relativity*. Oxford, UK: Oxford University Press, 2011.

[4] E. Poisson, A. Pound, and I. Vega, "The motion of point particles in curved spacetime," *Living Reviews in Relativity*, vol. 14, no. 7, 2011.

[5] S. E. Gralla and R. M. Wald, "A rigorous derivation of gravitational self-force," *Classical and Quantum Gravity*, vol. 28, no. 15, p. 159501, 2011.

[6] D. Puetzfeld, C. Lämmerzahl, and B. Schutz, eds., *Equations of motion in relativistic gravity*. Heidelberg, Germany: Springer, 2015.

[7] D. Malament, "A remark about the "geodesic principle" in general relativity," in *Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos* (M. Frappier, D. H. Brown, and R. DiSalle, eds.), pp. 245–252, New York: Springer, 2012.

[8] W. G. Dixon, "A covariant multipole formalism for extended test bodies in general relativity," *Il Nuovo Cimento*, vol. 34, no. 2, pp. 317–339, 1964.

[9] M. Mathisson, "Die mechanik des materieteilchens in der allgemeinen relativitätstheorie," *Zeitschrift für Physik*, vol. 67, pp. 826–844, 1931.

[10] J.-M. Souriau, "Modèle de particule à spin dans le champ électromagnétique et gravitationnel," *Annales de l'Institut Henri Poincaré Sec. A*, vol. 20, p. 315, 1974.

[11] A. Papapetrou, "Spinning Test Particles in General Relativity," /em Proc Roy Soc, vol. 290, pp. 248–258, 1951.

[12] S. Sternberg and V. Guillemin, *Symplectic Techniques in Physics*. Cambridge: Cambridge University Press, 1984.

[13] R. Geroch and P. S. Jang, "Motion of a body in general relativity," *Journal of Mathematical Physics*, vol. 16, no. 1, p. 65, 1975.

[14] J. Ehlers and R. Geroch, "Equation of motion of small bodies in relativity," *Annals of Physics*, vol. 309, pp. 232–236, 2004.

[15] J. O. Weatherall, "The motion of a body in Newtonian theories," *Journal of Mathematical Physics*, vol. 52, no. 3, p. 032502, 2011.