# A Remark About the "Geodesic Principle" in General Relativity*

Version 3.0

David B. Malament
Department of Logic and Philosophy of Science
3151 Social Science Plaza
University of California, Irvine
Irvine, CA 92697-5100
dmalamen@uci.edu

## 1 Introduction

General relativity incorporates a number of basic principles that correlate space-time structure with physical objects and processes. Among them is the

> Geodesic Principle: Free massive point particles traverse timelike geodesics.

One can think of it as a relativistic version of Newton's first law of motion.

It is often claimed that the geodesic principle can be recovered as a theorem in general relativity. Indeed, it is claimed that it is a consequence of Einstein's

*I am grateful to Robert Geroch for giving me the basic idea for the counterexample (proposition 3.2) that is the principal point of interest in this note. Thanks also to Harvey Brown, Erik Curiel, John Earman, David Garfinkle, John Manchak, Wayne Myrvold, John Norton, and Jim Weatherall for comments on an earlier draft.

equation (or of the conservation principle $\nabla_a T^{ab} = \mathbf{0}$ that is, itself, a consequence of that equation). These claims are certainly correct, but it may be worth drawing attention to one small qualification. Though the geodesic principle can be recovered as theorem in general relativity, it is not a consequence of Einstein's equation (or the conservation principle) *alone*. Other assumptions are needed to drive the theorems in question. One needs to put more in if one is to get the geodesic principle out. My goal in this short note is to make this claim precise (i.e., that other assumptions are needed).

All talk about deriving the geodesic principle is a bit delicate because it is not antecedently clear how to *formulate* it so that it is even a candidate for proof. One way or another, one has to confront the problem of how to associate an energy-momentum content $T_{ab}$ with a point particle. (Only then can one invoke the conservation principle $\nabla_a T^{ab} = \mathbf{0}$.) This is a problem even if one is willing to restrict attention to "test particles", i.e., even if one does not insist that $T_{ab}$ be recorded on the right side of Einstein's equation. One might try to work with energy-momentum "distributions" rather than proper smooth fields, but there is a natural alternative. In effect, one models a massive point particle as a nested sequence of small, but extended, bodies that converges to a point. One associates with each of the bodies a garden variety smooth energy-momentum field $T_{ab}$, and requires that, in each case, it satisfy certain constraints. Then one proves, if one can, that the point to which the bodies converge necessarily traverses a timelike geodesic.

Various theorems in the literature do, in fact, have this form. In all cases, one assumes that the energy-momentum field $T_{ab}$ associated with each small body in the sequence satisfies the conservation principle. (This captures the idea that the body is "free" i.e., not exchanging energy-momentum with some external field.) That much the theorems have in common. But they differ as to the additional constraints that are imposed. In some cases, very specific assumptions are made about the constitution of the bodies in the sequence. A theorem in Thomas [6] and Taub [5] is of this type. There one takes each body to be a blob of perfect fluid, with everywhere non-negative isotropic pressure, that satisfies a strong constraint. It is required that the pressure at every point in the blob remains constant over time. Given this assumption (and the conservation principle), it is easy to prove that the convergence point of the bodies does, in fact, traverse a timelike geodesic.

This result is certainly of interest. But it seems a considerable advance to prove theorems that dispense with special modeling assumptions in favor of generic ones. The result of Geroch and Jang [2] that I'll formulate in section 3 (proposition 3.1) is an example of this latter type. There one only assumes that the energy-momentum field $T_{ab}$ of each body in the sequence satisfies a certain "energy condition". It asserts, in effect, that, whatever else is the case, energy propagates within the body at velocities that are timelike. That too is sufficient, together with the conservation principle, to guarantee that the convergence point of the bodies traverses a timelike geodesic.

My point in this note is that the Geroch-Jang theorem fails if one drops the energy-condition requirement. As we shall see (proposition 3.2), *the conservation condition alone imposes no restrictions whatsoever on the wordline of the convergence point of the bodies.* It can be a null or spacelike curve. It can also be a timelike curve that exhibits any desired pattern of large and/or changing acceleration.

In the Geroch-Jang theorem, one allows oneself to ignore the negligible effect on the background metric made by (the energy-momentum content of) each body in the convergent sequence. A stronger result of Ehlers and Geroch [1] relaxes this restriction. There it is not required that the perturbative effect disappear entirely at each intermediate stage, but only that, in a certain precise sense, it disappear in the limit. In this result too, an energy condition is imposed in lieu of any more specific modeling assumptions about the bodies in the sequence. And again in this case, the result fails completely without the energy condition. (The counterexample that we present for the weaker theorem (in proposition 3.2) carries over intact to the stronger one.) To keep the presentation as simple as possible, I will limit my attention to the former.

## 2 The Energy-Momentum Field $T_{ab}$

In this section, we review a few things about the energy-momentum field $T_{ab}$ that will be important later.[1] Some readers may want to skip to section 3.[2]

In what follows, let $(M, g_{ab})$ be a relativistic spacetime, which we here take to consist of a smooth, connected, four-dimensional differential manifold $M$, and a smooth metric $g_{ab}$ on $M$ of Lorentz signature $(1, 3)$. With this sign convention, a vector $\xi^a$ at a point counts as timelike if $\xi^a \xi_a > 0$, null if $\xi^a \xi_a = 0$, causal if $\xi^a \xi_a \geq 0$, and spacelike if $\xi^a \xi_a < 0$. We assume that $(M, g_{ab})$ is temporally orientable, and that some temporal orientation has been specified.

Let us start with point particles. It is a basic assumption of relativity theory that we can associate with every point particle, at every point on its worldline, a *four-momentum* (or *energy-momentum*) vector $P^a$ that is tangent to its worldline. We can think of it as encoding several pieces of information. It is standardly taken for granted that $P^a$ is causal. In that case, at least, the length of $P^a$ gives the *mass* of the particle:

$$\text{mass} = (P^a P_a)^{\frac{1}{2}}.$$

So, in particular, the mass of the particle is strictly positive iff its four-momentum vector field is timelike. Let $\xi^a$ be a future-directed, unit timelike vector at some point on the worldline of the particle. We can think of it as representing the instantaneous state of motion of a background observer at that point. Suppose we decompose $P^a$ into two component vectors that are, respectively, proportional to, and orthogonal to, $\xi^a$:

$$P^a = \underbrace{(P^b \xi_b)}_{\text{energy}} \xi^a + \underbrace{(P^a - (P^b \xi_b)\, \xi^a)}_{3-\text{momentum}}. \tag{2.1}$$

The proportionality factor $P^b \xi_b$ in the first is standardly understood to give the *energy* of the particle relative to $\xi^a$; and the second component is understood to give the *three-momentum* of the particle relative to $\xi^a$.

[1] We will assume familiarity with the basic mathematical formalism of general relativity in what follows. For background material, see, e.g., Hawking and Ellis [3], Wald [7], or Malament [4]. The third is a set of unpublished lecture notes that is available online.

[2] All the material in the section is perfectly standard except for one small bit of *ad hoc* terminology. In addition to the weak and dominant energy conditions, we will consider something that we call the "strengthened dominant energy condition".

Let us now switch from point particles to matter fields, e.g., fluids and electromagnetic fields. Each such field is represented by one or more smooth tensor (or spinor) fields on the spacetime manifold $M$. Each is assumed to satisfy field equations involving the spacetime metric $g_{ab}$.

For present purposes, the most important basic assumption about the matter fields is the following.

> Associated with each matter field $\mathcal{F}$ is a symmetric smooth tensor field $T_{ab}$ characterized by the property that, for all points $p$ in $M$, and all future-directed, unit timelike vectors $\xi^a$ at $p$, $T^a{}_b\,\xi^b$ is the four-momentum density of $\mathcal{F}$ at $p$ as determined relative to $\xi^a$.

$T_{ab}$ is called the *energy-momentum* field associated with $\mathcal{F}$. The four-momentum density vector $T^a{}_b\xi^b$ at $p$ can be further decomposed into components proportional to, and orthogonal to, $\xi^a$ (just as with the four-momentum vector $P^a$):

$$T^a{}_b\,\xi^b = \underbrace{(T_{nb}\,\xi^n\xi^b)}_{\text{energy density}}\,\xi^a + \underbrace{(T^a{}_b\,\xi^b - (T_{nb}\,\xi^n\xi^b)\,\xi^a)}_{3-\text{momentum density}}. \tag{2.2}$$

The coefficient of $\xi^a$ in the first component, $T_{ab}\xi^a\xi^b$, is the *energy density* of $\mathcal{F}$ at $p$ as determined relative to $\xi^a$. The second component, $T_{nb}(g^{an} - \xi^a\,\xi^n)\xi^b$, is the *three-momentum density* of $\mathcal{F}$ at $p$ as determined relative to $\xi^a$.

Various assumptions about matter fields can be captured as constraints on the energy-momentum tensor fields with which they are associated. The Geroch-Jang theorem makes reference to the third and fourth in the following list. (Suppose $T_{ab}$ is associated with matter field $\mathcal{F}$.)

**Weak Energy Condition:** For all points $p$ in $M$, and all unit timelike vectors $\xi^a$ at $p$, $T_{ab}\,\xi^a\xi^b \geq 0$.

**Dominant Energy Condition**: For all points $p$ in $M$, and all unit timelike vectors $\xi^a$ at $p$, $T_{ab}\,\xi^a\xi^b \geq 0$ and $T^a{}_b\,\xi^b$ is causal.

**Strengthened Dominant Energy Condition**[3]: For all points $p$ in $M$, and all unit timelike vectors $\xi^a$ at $p$, $T_{ab}\,\xi^a\xi^b \geq 0$ and, if $T_{ab} \neq \mathbf{0}$, then $T^a{}_b\,\xi^b$ is timelike.

[3] This is not a standard name.

**Conservation Condition**: $\nabla_a T^{ab} = \mathbf{0}$ at all points in M.

The weak energy condition asserts that the energy density of $\mathcal{F}$ (as determined relative to any background observer) is everywhere non-negative. The dominant energy condition adds the requirement that the energy-momentum density of $\mathcal{F}$ (as determined relative to a background observer) is causal. It can be understood to assert that the energy of $\mathcal{F}$ does not propagate at superluminal velocity (relative to any such observer). The strengthened version of the condition just changes "causal" to "timelike". Each of the energy conditions is strictly stronger than the ones that precede it.[4]

The final condition in the list captures the requirement that the energy-momentum carried by $\mathcal{F}$ be locally conserved. If two or more matter fields are present in the same region of spacetime, it need not be the case that each one individually satisfies the condition. Interaction may occur. But presumably in that case the composite energy-momentum field formed by taking the sum of the individual ones satisfies the condition. Energy-momentum can be transferred from one matter field to another, but it cannot be created or destroyed.

Suppose $T_{ab}$ represents the aggregate energy-momentum present in some region of spacetime. Then, at least if it is understood to arise from "source fields" rather than "test fields", it must satisfy Einstein's equation

$$R_{ab} - \frac{1}{2}\, R\, g_{ab} \;=\; 8\,\pi\, T_{ab}.$$

The left side is divergence-free: $\nabla_a (R^{ab} - \frac{1}{2}\, R\, g^{ab}) \;=\; \mathbf{0}$. (This follows from Bianchi's identity.) So, in this (source field) case at least, the conservation condition is a consequence of Einstein's equation.

The dominant energy and conservation conditions have a number of joint consequences that support the interpretations just given. Here is one. It requires a preliminary definition.

Let $(M,\, g_{ab})$ be a fixed relativistic spacetime, and let $S$ be an achronal subset of $M$ (i.e., a subset no two points of which are connected by a smooth timelike curve). The *domain of dependence* $D(S)$ of $S$ is the set of all points $p$ in $M$

[4] If $\lambda^a$ is a smooth spacelike field, then $T_{ab} = \lambda_a \lambda_b$ satisfies the weak, but not the dominant, energy condition. Similarly, if $\lambda^a$ is a smooth, non-vanishing null field, then $T_{ab} = \lambda_a \lambda_b$ satisfies the dominant, but not the strengthened dominant, energy condition.

with this property: given any smooth causal curve without (past or future) endpoint,[5] if (its image) passes through $p$, then it necessarily intersects $S$.

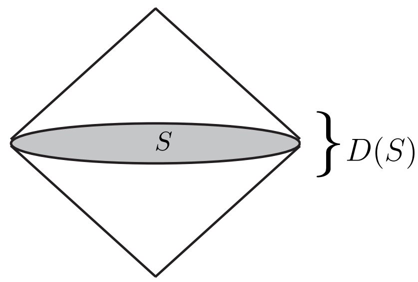


Figure 2.1: The domain of dependence $D(S)$ of an achronal set $S$.

**Proposition 2.1.** *Let $S$ be an achronal subset of $M$. Further let $T_{ab}$ be a smooth symmetric field on $M$ that satisfies both the dominant energy and conservation conditions. Finally, assume $T_{ab} = \mathbf{0}$ on $S$. Then $T_{ab} = \mathbf{0}$ on all of $D(S)$.*

The intended interpretation of the proposition is clear. If energy-momentum cannot propagate (locally) outside the null-cone, and if it is conserved, and if it vanishes on $S$, then it must vanish throughout $D(S)$. After all, how could it "get to" any point in $D(S)$? Note that our formulation of the proposition does not presuppose any particular physical interpretation of the symmetric field $T_{ab}$. All that is required is that it satisfy the two stated conditions. (For a proof, see Hawking and Ellis [3, p. 94].)

# 3 A Theorem and A Counterexample

Now we turn to the Geroch-Jang theorem [2] itself.

**Proposition 3.1.** *Let $(M, g_{ab})$ be a relativistic spacetime, and let $\gamma : I \to M$ be a smooth curve. Suppose that given any open subset $O$ of $M$ containing $\gamma[I]$, there exists a smooth symmetric field $T_{ab}$ on $M$ such that:*

[5]Let $\gamma : I \to M$ be a smooth curve. We say that a point $p$ in $M$ is a *future-endpoint* of $\gamma$ if, for all open sets $O$ containing $p$, there exists an $s_0$ in $I$ such that for all $s \in I$, if $s \geq s_0$, then $\gamma(s) \in O$, i.e. the image of $\gamma$ eventually enters and remains in $O$. (*Past-endpoints* are defined similarly.)

*(1)* $T_{ab}$ *satisfies the strengthened dominant energy condition;*

*(2)* $T_{ab}$ *satisfies the conservation condition;*

*(3)* $T_{ab} = \mathbf{0}$ *outside of* $O$*;*

*(4)* $T_{ab} \neq \mathbf{0}$ *at some point in* $O$*.*

*Then* $\gamma$ *is a timelike curve, and can be reparametrized so as to be a geodesic.*

The proposition might be paraphrased this way. Suppose that arbitrarily small bodies (with energy-momentum) satisfying conditions (1) and (2) can contain the image of a curve $\gamma$ in their worldtubes. Then $\gamma$ must be a timelike geodesic (up to reparametrization). In effect, as discussed above, we are representing "point particles" as nested convergent sequences of smaller and smaller extended bodies. Bodies here are understood to be "free" if their internal energy-momentum is conserved (by itself). If a body is acted upon by a field, it is only the composite energy-momentum of the body and field together that is conserved.

The proof proceeds by showing that given any worldtube and any energy-momentum field satisfying conditions (1)-(4), the tube must contain the image of a timelike geodesic. That cannot be true for arbitrarily small tubes containing the image of the original curve $\gamma$ unless that curve itself is a timelike geodesic (up to reparametrization).

Our formulation of the proposition takes for granted that we can keep the background spacetime metric $g_{ab}$ fixed while altering the fields $T_{ab}$ that live on $M$. This is justifiable only to the extent that, once again, we are dealing with test bodies whose effect on the background spacetime structure is negligible.

Though, of course, the proposition has an intended interpretation, it is important that it stands on its own as a well-formed mathematical theorem (as does proposition 2.1). It can be proved without any appeal to the interpretation of $T_{ab}$. It is also noteworthy in the proposition that we do not have to *assume* that the initial curve $\gamma$ is timelike. That is something that we *prove*.

Our main claim, as announced above, is that the proposition fails if condition (1) is dropped. Without it, one cannot prove that the original curve $\gamma$ must be a geodesic (up to a reparametrization), not even if we *do* assume in advance that it is timelike. The following proposition gives a counterexample.

Figure 3.1: A non-geodesic timelike curve enclosed in a tube (as considered in proposition 3.2).

**Proposition 3.2.** *Let* $(M, g_{ab})$ *be Minkowski spacetime, and let* $\gamma : I \to M$ *be* any *smooth timelike curve. Then given any open subset* $O$ *of* $M$ *containing* $\gamma[I]$*, there exists a smooth symmetric field* $T_{ab}$ *on* $M$ *that satisfies conditions (2), (3), and (4) in the preceding proposition. (If we want, we can also strengthen condition (4) and require that* $T_{ab}$ *be non-vanishing throughout some open subset* $O_1 \subseteq O$ *containing* $\gamma[I]$*.)*

*Proof.* Let $O$ be an open subset of $M$ containing $\gamma[I]$, and let $f : M \to \mathbb{R}$ be *any* smooth scalar field on $M$. (Later we will impose further restrictions on $f$.) Consider the fields $S^{abcd} = f(g^{ad}g^{bc} - g^{ac}g^{bd})$ and $T^{ac} = \nabla_b \nabla_d S^{abcd}$, where $\nabla$ is the (flat) derivative operator on $M$ compatible with $g_{ab}$. (So $\nabla_a g_{bc} = \nabla_a g^{bc} = \mathbf{0}$.) We have

$$T^{ac} = (g^{ad}g^{bc} - g^{ac}g^{bd}) \nabla_b \nabla_d f = \nabla^c \nabla^a f - g^{ac} (\nabla_b \nabla^b f). \tag{3.1}$$

So $T^{ac}$ is clearly symmetric. It is also divergence-free since

$$\nabla_a T^{ac} = \nabla_a \nabla^c \nabla^a f - \nabla^c \nabla_b \nabla^b f = \nabla^c \nabla_a \nabla^a f - \nabla^c \nabla_b \nabla^b f = \mathbf{0}.$$

(The second equality follows from the fact that $\nabla$ is flat, and so $\nabla_a$ and $\nabla^c$ commute in their action on arbitrary tensor fields.)

To complete the proof, we now impose further restrictions on $f$ to insure that conditions (3) and (4) are satisfied. Let $O_1$ be any open subset of $M$ such that

$\gamma[I] \subseteq O_1$ and $\mathrm{cl}(O_1) \subseteq O$. (Here $\mathrm{cl}(A)$ is the closure of $A$.) Our strategy will be to choose a particular $f$ on $O_1$, and a particular $f$ on $M - \mathrm{cl}(O)$, and then fill-in the buffer zone $\mathrm{cl}(O) - O_1$ any way whatsoever (so long as the resultant field is smooth). On $M - \mathrm{cl}(O)$, we simply take $f = 0$. This choice guarantees that, no matter how we smoothly extend $f$ to all of $M$, $T^{ac}$ will vanish outside of $O$.

For the other specification, let $p$ be any point in $M$, and let $\chi^a$ be the "position field" on $M$ determined relative to $p$. So $\nabla_a \chi^b = \delta_a{}^b$ everywhere, and $\chi^a = \mathbf{0}$ at $p$. (See, for example, proposition 1.7.11 in Malament [4].) On $O_1$, we take $f = -(\chi^n \chi_n)$. With that choice, $T^{ac}$ is non-vanishing at all points in $O_1$. Indeed, we have

$$\nabla_a f \;=\; -2\,\chi_n \nabla_a \chi^n \;=\; -2\,\chi_n\,\delta_a{}^n \;=\; -2\,\chi_a,$$

and, therefore,

$$\begin{aligned} T^{ac} \;&=\; \nabla^c\,\nabla^a f \,-\, g^{ac}\,(\nabla_b\,\nabla^b f) \;=\; -2\,\nabla^c \chi^a \,+\, 2\,g^{ac}\,(\nabla_b\,\chi^b) \\ &=\; -2\,g^{ca} + 2\;g^{ac}\,\delta_b{}^b \;=\; -2\,g^{ac} \,+\, 8\,g^{ac} \;=\; 6\,g^{ac} \end{aligned}$$

throughout $O_1$. □

One point about the proof deserves comment. As restricted to $O_1$ and to $M - \mathrm{cl}(O)$, the field $T_{ab}$ that we construct *does* satisfy the strengthened dominant energy condition. (In the first case, $T_{ab} = 6\,g_{ab}$, and in the second case, $T_{ab} = \mathbf{0}$.) But we know – from the Geroch-Jang theorem itself – that it cannot satisfy that condition everywhere. So it must fail to do so in the buffer zone $\mathrm{cl}(O) - O_1$. That shows us something. We can certainly choose $f$ in the zone so that it smoothly joins with our choices for $f$ on $O_1$ and $M - \mathrm{cl}(O)$. But, no matter how clever we are, we cannot do so in such a way that $T^{ab}$ (as expressed in (3.1)) satisfies the strengthened dominant energy condition.

## References

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

[2] R. Geroch and P. S. Jang. Motion of a body in general relativity. *Journal of Mathematical Physics*, 16:65–67, 1975.

[3] S.W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge University Press, 1972.

[4] D. Malament. *Lecture Notes on the Foundations of Relativity Theory*. http://www.lps.uci.edu/malament/FndsofGR.html, unpublished.

[5] A. H. Taub. On Thomas' result concerning the geodesic hypothesis. *Proceedings of the National Academy of Sciences*, 48:1570–1571, 1962.

[6] T. Y. Thomas. On the geodesic hypothesis in the theory of gravitation. *Proceedings of the National Academy of Sciences*, 48:1567–1569, 1962.

[7] R. Wald. *General Relativity*. University of Chicago Press, 1984.