

Introduction and teaching of relativity theory

P.G.L. Porta Mana 

Western Norway University of Applied Sciences [<pgl@portamana.org>](mailto:pgl@portamana.org)

Draft. 25 March 2023;; updated 8 April 2023

Notes on a possible way to introduce and teach general, Lorentzian, and Newtonian relativity.

1 Motivation

The way Lorentzian and general relativity are introduced in many books leaves some uneasiness as regards some logical steps in the underlying reasoning – although I feel no doubt about the internal logical consistency of the results. For example, if we come to the conclusion that metre sticks shorten or are seen shortened by some observer, then is our initial reasoning about some experiments, such as the Michelson-Morley one, still logically valid? How were those meters initially defined? If general relativity makes the notion of rigid body very tricky, then where does our initial reasoning involving rigid rods logically stand?

The present note tries to develop the theory from some postulates and notions that can be induced from experiments – at least from thought experiments – while avoiding logical loop-holes. It is only tentative, a step, probably still affected by inconsistencies.

The presentation is, for the moment, mostly in form of main points and self-reminders.

2 Initial notions and postulates

We start with the notion of “event”, something that has very small extension in space and time. Intuitive, qualitative notions of “extension” should suffice for such a definition. Spacetime is introduced as the four-dimensional topological space made of such events. For the moment the only thing we can say about events is whether they are coincident (or very close, in a qualitative sense) or not.

Next we introduce the notion of a persistent clock. “Persistent” in the sense that it is not just an event, but rather a continuous sequence of

events, which has a very small extension in space. In spacetime it can be represented by a one-dimensional curve, which is called a “worldline”. “Clock” in the sense that there is some kind of periodicity in this sequence of events. Note how tricky the notion of “periodicity” is, though: how do we establish that some phenomenon is periodic, without having a notion of clock? For the moment we must thus take “periodic” as a primitive notion. It’s the refinement of our own biological feeling of “equal time intervals”. We call such a persistent clock an “observer”.

Next we make an experimentally verifiable assumption. Whenever two observers coincide – they have coincident worldlines – and their clocks are of the same making, then they notice and exact equality of the periods of the two clocks. The absolute times shown by the clocks – the number of periods counted from some initial moment – may be different, but the period intervals are the same.

Let us now imagine that two such observers are coincident in some initial and final parts of their worldlines, but not in all their internal parts. Let us also assume that in their initial, coincident part, their clocks are synchronized, in the sense that they give the same absolute time, not just same time intervals. Experimentally it is observed that if the worldlines diverge and then come to coincide again, at the point of reunion the two clocks will have generally different absolute times (though both future times with respect to the initial point of worldline separation), but they have again same time intervals in this new coincident part of the curve. Generalizing to three or more observers, in general we would find three or more different absolute times at the point of reunion.

This experimental fact is at variance with Newtonian relativity, which instead assumes that in such a situation the two clocks would still have identical absolute times upon reuniting.

We also take as primitive the possibility of sending light signals from a given observer, using some kind of device.

Another experimentally verifiable assumption is that if two coincident observers send a light signal in the same direction, using light-sending devices of the same making, then the two signals also have coincident worldlines.

This experimental assumption is extended, however. Suppose we have two observers with non-coincident worldlines, which however intersect in one event. At that event, both observers send light signals

in the same direction. Then also in this case the two light signals have coincident worldlines. This is an experimentally verifiable assumption.

This last assumption is extremely important because it selects, in spacetime, a special set of worldlines and of “lightcones”. A lightcone is the set of worldlines of light signals stemming from the same event but having different directions.

The last assumption could have been restated as saying that the velocity of a light signal is independent of the velocity of its source. But we have thus far happily avoided the tricky notions of “velocity” and “distance”.

3 Remarks on the differential-geometric representation

To each event we can also associate scalar values representing measurement results, thus obtaining scalar-valued functions over regions of the spacetime. We can also consider functions from a one-dimensional scalar space into spacetime. Spacetime with its collection of possible functions from and into can thus be represented by a differential manifold. The functions are so chosen that, if we “zoom in” on a region around an event, the level hypersurfaces of the functions from the manifold look like hyperplanes (or a constant value over the whole region), and the images of the functions onto the manifold look like straight lines (or points).

This also implies that if we consider a tangent vector v at the event P , we can approximately associate another event close to P with the vector ϵv , if ϵ is small enough. The event approximately associated with ϵv is the one having coordinates $x^i(P) + \epsilon dx^i(P) \cdot v + O(\epsilon^2)$ in *any* coordinate system (x^i) admitted by the differential manifold. This is possible because coordinate transformations only affect $O(\epsilon^2)$ terms.

4 Metric and radar coordinates

The metric (field) g encodes the length of physical time intervals for all observers. Given the worldline $s \mapsto C(s)$ of an observer between events $C(s_0)$ and $C(s_1)$, the physical time elapsed for that observer between those two events is given by

$$\int_{s_0}^{s_1} \frac{1}{c} \sqrt{|\dot{C}(s) g[C(s)] \dot{C}(s)|} ds . \quad (1)$$

If v is the tangent vector at some event to a worldline, then the physical time interval between that event and one roughly corresponding to a parameter increment Δs along the worldline is $\frac{1}{c}\sqrt{|v g v|} \Delta s + O(\Delta s^2)$, where g is the metric at the first event. We can normalize v in such a way that the increment Δs itself is the physical time interval; this corresponds to a local change of the worldline parameter s . With such normalization we must then have

$$|v g v| = c^2 \quad \text{if } v \text{ is normalized,} \quad (2)$$

and we call v the local *4-velocity*. Note that this normalization assumes that the metric has dimensions of a squared-length, and that the 4-velocity has dimensions of inverse-time. Any vector u tangent to a worldline can be normalized by multiplying it by $c/\sqrt{|u g u|}$. Such multiplication also gives it dimensions of inverse time.

Given the 4-velocity v of an observer at an event P , there's a special three-dimensional set of events in the spatial and temporal vicinity of P . Each event in this set has the property that a light signal sent by the observer a small interval of time Δs before P , and bouncing off the event, reaches back the observer an interval Δs after P , except for differences $O(\Delta s^2)$. Such events are said to be locally orthogonal to the observer's 4-velocity v at P . The observer can conventionally consider these local neighbouring events as *simultaneous* to P .

To each such orthogonal or simultaneous neighbouring event, the observer can also associate a distance, by convention. The distance is defined in such a way that the "ordinary" velocity of the signal reaching the event is equal to c . The distance must therefore be defined as $\Delta s/c$.

The constructions above can be made with respect to events on the worldline of the observer very close to P . This leads to a local set of coordinates in the neighbourhood of P called *radar coordinates*. The observer can therefore conventionally associate a "time" and a "distance", and thus a "velocity", to all events or worldlines in the neighbourhood of P .

We can associate a covector

$$\underline{v} := \frac{v g}{v g v} \equiv -\frac{1}{c^2} v g \quad (3)$$

to a 4-velocity v , defined by the following properties: (a) contracted with v it yields unity: $\underline{v} v = 1$, (b) it yields zero when applied to any vector orthogonal to v . We call it the '4-covelocity'.

The 4-velocity and 4-covelocity allow us to associate a conventional radar-distance and radar-time to events in the neighbourhood of an event P on a worldline with 4-velocity \underline{v} at P . For a neighbouring event approximately identified by the small tangent vector $\epsilon \underline{u}$, its time with respect to P is $\epsilon \underline{v} \underline{u}$, and its position vector is $\epsilon (\underline{I} - \underline{v} \otimes \underline{v}) \underline{u}$. If another worldline in the neighbourhood of P has tangent vector \underline{u} , then its conventional radar-coordinate velocity V from the point of view of the observer with 4-velocity \underline{u} is

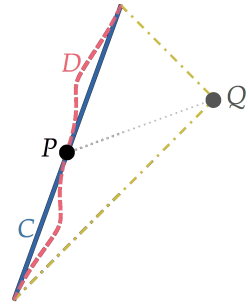
$$V := \frac{(\underline{I} - \underline{v} \otimes \underline{v}) \underline{u}}{\underline{v} \underline{u}} \equiv \frac{\underline{u} - \underline{v} (\underline{v} \underline{u})}{\underline{v} \underline{u}}. \quad (4)$$

This expression is valid even if \underline{u} is not normalized. In radar coordinates with respect to \underline{v} , the 4-velocity of the second worldline has components

$$\left(1 - \frac{V^2}{c^2}\right)^{-\frac{1}{2}} (1, V). \quad (5)$$

These formulae also reveal the “time dilation” and “length contraction” phenomena. A physical-time lapse of ϵ for the observer with 4-velocity \underline{u} has a larger *radar-time lapse* of $\epsilon \underline{v} \underline{u} = \epsilon / (1 - V^2/c^2)$ according to \underline{v} , since $1 - V^2/c^2 \leq 1$. That is, observer \underline{v} attributes a *conventional* time lapse larger than ϵ to a *physical* time lapse of ϵ on the clock of \underline{u} , which therefore “runs slowly” according to the conventional time reckoning of \underline{v} . It should again be emphasized that the radar-time lapse is just a conventional time coordinate, not a physical one. Also note that radar-time is not the time at which \underline{v} *sees* other events.

The extent of the neighbourhood around an event P in which radar distances and velocities are well-defined depends on the problem considered; in general it cannot be extended indefinitely. It is important that the worldline of the observer measuring the radar distance can be considered straight, otherwise the result will not make much sense. In the figure on the right, for example, the observer with worldline D (dashed red)



than the observer with worldline C (solid blue) would, even if both observers have the same tangent vectors at P and at the events in which the light signals (dot-dashed yellow) are sent and received.

Other kinds of “distances” and “velocities” can be *defined* in more extended regions. But it is important to keep in mind that all such definitions have an element of arbitrariness, and Newtonian intuition should not be applied in such situations. Speaking of the “distances” and “velocity” of faraway galaxies, for example, is a subtle matter relying on many conventions and prone to misunderstandings¹. It would probably be much better if less intuitive and less misleading terms were used instead.

✍ Notion of “constant distance or position” with respect to an observer

✍ Extensibility of coordinate systems

5 Observer decomposition of geometric objects

Most physico-geometric objects of interest in general relativity are the tensor product of a multivector and a multivector and a multivector. They can be called multi(co)vector-valued multivectors. The multivector part usually represent the possibility of locally integrating the object over a small segment, area, volume, or 4-volume.

An *instantaneous* observer at some event P with instantaneous 4-velocity u can always decompose such an object into a “time-aligned” and “space-aligned” part. The time-aligned part of an m -vector is an $m-1$ -vector, and the space-aligned part another m -vector. Given a metric and consequent notion of instantaneous simultaneous space to the observer, the time-aligned and space-aligned parts have the property of being orthogonal to the 4-velocity u ; hence the observer can interpret both as “spatial” multivectors. An analogous situation holds for multivectors.

Balance laws are expressed as the vanishing of the exterior derivative or of the exterior covariant derivative of such objects. These derivatives can also be decomposed into time-aligned and space-aligned parts of the time- and space-aligned components of the original object. Such decomposition, however, *depends on how the instantaneous observer is extended in the neighbouring region of spacetime.*



¹ see e.g. Davis et al. 2003; Davis & Lineweaver 2004; Grøn 2004; Weber 2004.

6 Energy, momentum, stress, and amount of matter

Before discussing the notions of amount of matter, force, momentum, energy, it is best to clarify and emphasize important differences, but also important similarities, between Newtonian and Lorentzian relativity.

A fundamental primitive idea in Newtonian and Lorentzian relativity is that of something that persists and we can follow through time; something that preserves its identity through time. The word “thing” indeed is already associated with such a primitive idea. It is implicit in the notions of “body”, “particle”, “matter”. This primitive idea was so fundamental in the history of physics that any other physical notions were for a long time translated either in “things” themselves or in movement of such “things”. This is very evident in the history of the notions of light and aether², and in the slow acceptance of the notion of *field* as a sort of “thing” that we associate with a point in time and space but doesn’t have a trajectory. We shall see later that the primitive idea of “identity” or “persistence” can be approached from another point of view, from which it also shows more similarities with the idea of field.

In Newtonian relativity we associate matter and mass, and distinguish them from energy. In Lorentzian and general relativity we have to dissociate the notions of matter and mass instead, and almost identify the latter with energy. Luckily this process can partly be already achieved in Newtonian relativity.

✂ Start with balance of amount of matter (“moles”, stoichiometry). In Newtonian relativity, balance of mass is a consequence of this.

Introduce 3-covector fields: they naturally represent “bodies” and things having trajectories.

In both Newtonian and Lorentzian relativity we can introduce a matter twisted 3-covector field N . Given a small spacelike 3-volume, any observer can trace a world-tube in spacetime out of it. A worldline is just such world-tube, idealized to shrink to a vanishing spatial cross-section.

Balance of amount of matter is expressed by the equation

$$dN = 0 . \quad (6)$$

At a spacetime event we can choose an observer with 4-velocity u such that³

$$u \cdot N = 0 , \quad u g u = -c^2 ; \quad (7)$$

² Whittaker 1951; 1953. ³ cf. Eckart 1940 p. 920.

the first equation says that \underline{u} lies in the kernel of N . This 4-velocity is explicitly given by

$$\underline{u} := \frac{c \gamma^{-1} \cdot N}{\sqrt{|(\gamma^{-1} \cdot N) \mathbf{g} (\gamma^{-1} \cdot N)|}} . \quad (8)$$

The exterior covariant derivative is denoted “ D ”.

Balance of 4-momentum is expressed by the equation

$$D\mathcal{T} = 0 . \quad (9)$$


We express the 4-momentum associated to matter as

$$\mathcal{T} = e \underline{u} \otimes N + \underline{u} \otimes \underline{q} + \underline{p} \otimes N + \sigma \wedge \underline{u} \quad (10)$$

with the properties

$$\underline{u} \mathbf{g} \underline{p} = 0 , \quad \underline{u} \mathbf{g} \sigma = 0 , \quad \underline{q} \mathbf{g}^3 N = 0 , \quad (11)$$

where e is a scalar representing molar energy density, \underline{p} is a 1-covector representing molar momentum density, \underline{q} a 2-covector representing areic heat flux, and σ a 1-covector-valued 2-covector representing compressive stress, or pressure for short.

 The fact that matter has energy and inertia comes from the fact that the energy-momentum-stress tensor depends, among other quantities, on N .

Appendices

A Geometric remarks and conventions

I summarize here some remarks and convenient conventions of a geometric character.

A.1 Coordinates and orderings

Coordinates in a region are typically denoted t, x, y, z , corresponding to indices $0, 1, 2, 3$. Coordinate t has dimensions of time, x, y, z of length. With these coordinates we associate an inner orientation in the given order.

To work with multivector and multicovector spaces and with orientations, It is useful to set up the following correspondence between various

ordered groupings of coordinates (the symbol “ $\hat{=}$ ” means “corresponds to”):

$$\begin{aligned} 1 &\hat{=} txyz \\ t &\hat{=} xyz & x &\hat{=} tzy & y &\hat{=} txz & z &\hat{=} tyx \\ tx &\hat{=} yz & ty &\hat{=} zx & tz &\hat{=} xy \end{aligned} \quad (12)$$

the correspondence being symmetric.

We take the orderings given above to be also the standard orientations in corresponding subspaces. For example, a 3-space extending in the t, y, z directions, that is, with a value $x = \text{const}$, has standard orientation tyz .

A.2 Bases of multivector and multicovector spaces

The coordinate system induces the basis dt, dx, dy, dz in each tangent covector space and the dual $\partial_t, \partial_x, \partial_y, \partial_z$ in each tangent vector space.

From two vectors such as ∂_x and ∂_y we can form a new 2-dimensional geometric object: an area of magnitude equal to that of the parallelogram formed by the two vectors but whose shape is unimportant, residing on the plane spanned by them, and inner-oriented in the direction defined by ∂_x followed by ∂_y . We call this a 2-vector and denote its construction by the antisymmetric product $\partial_x \wedge \partial_y$. We can generalize this construction to multivectors and multicovectors. See Burke (1987; 1995; 1983), Bossavit (1991), Schouten (1989) for more details and a geometric understanding of these objects [for the moment](#).

Let us use the abbreviations

$$d^2xy := dx \wedge dy, \quad \partial_{xy}^2 := \partial_x \wedge \partial_y, \quad (13)$$

and so on for wedge products of basis vectors and covectors.

We choose the following ordered basis and dual on 2-covector and 2-vector spaces:

$$\begin{aligned} d^2tx, \quad d^2ty, \quad d^2tz, \quad d^2yz, \quad d^2zx, \quad d^2xy; \\ \partial_{tx}^2, \quad \partial_{ty}^2, \quad \partial_{tz}^2, \quad \partial_{yz}^2, \quad \partial_{zx}^2, \quad \partial_{xy}^2. \end{aligned} \quad (14)$$

For 3-covector and 3-vector spaces we choose

$$\begin{aligned} d^3xyz, \quad d^3tzy, \quad d^3txz, \quad d^3tyx; \\ \partial_{xyz}^3, \quad \partial_{tzy}^3, \quad \partial_{txz}^3, \quad \partial_{tyx}^3. \end{aligned} \quad (15)$$

The 4-vector and 4-covector spaces are one-dimensional, with bases d^4txyz and ∂^4_{txyz} .

A.3 Metrics on multivector and multicovector spaces

A metric \mathbf{g} on the tangent 1-vector space also induces metrics on multivector spaces and multicovector spaces. Let us denote $\overset{3}{\mathbf{g}}$ the matrix on the 2-vector space. It is a 3-covector-valued 3-vector whose action is defined by

$$(\mathbf{u} \wedge \mathbf{u}') \overset{3}{\mathbf{g}} (\mathbf{v} \wedge \mathbf{v}') = (\mathbf{u} \wedge \mathbf{u}') \cdot [(\mathbf{g}\mathbf{v}) \wedge (\mathbf{g}\mathbf{v}')]. \quad (16)$$

This generalizes to multivectors, and to multicovectors using the inverse metric \mathbf{g}^{-1} .

The general rule is that the metric product between a multicovector basis element with ordered coordinates $I := ii'i'' \dots$ and one with coordinates $J := jj'j'' \dots$ is given by the minor $\det(g_{I,J})$ obtained by taking rows I and columns J from the matrix representing \mathbf{g} . Analogously for multicovectors and minors of the inverse matrix representing \mathbf{g}^{-1} .

With our choice of bases (14), (15) and the correspondence (12) we have various convenient equalities involving these metrics, for example

$$\begin{aligned} \partial^3_{xyz} \overset{3}{\mathbf{g}} \partial^3_{tzy} &= g^{tx} \det \mathbf{g}, \\ d^3txz \overset{3}{\mathbf{g}} \partial^3_{tyx} &= g_{yz} \det \mathbf{g}, \\ \partial^3_{tx} \overset{2}{\mathbf{g}} \partial^3_{zx} &= \det(g_{tx,zx}) = \det(g^{yz,ty}) \det \mathbf{g}, \end{aligned} \quad (17)$$

without the appearance of minus signs.

A.4 Inner- and outer-oriented objects



Given the ordered coordinate system t, x, y, z on a particular region, we define there the twisted unit function $P \mapsto \tilde{1}$ (which can be extended to the whole manifold only if the latter is inner-orientable). Note that $\tilde{1} \tilde{1} = 1$, the unit function.

The symbol “ \sim ” denotes multiplication by the twisted unit function, and transforms inner-oriented objects into outer-oriented ones and vice versa. When an untwisted object with orientation o is multiplied by the

twisted unit function, it acquires a twisted orientation \tilde{o} given by the rule

$$\text{or}(\tilde{o}, o) = \text{or}(txyz) . \quad (18)$$

For example, $\text{or}(d^3 \tilde{x}yz) = \text{or}(t)$ and $\text{or}(\partial_{ty}^2) = \text{or}(zx)$. Note the appearance of the correspondences (12).

In order to keep track of the orientation of twisted objects, a notation according to the following examples is convenient:

$$d^3 \tilde{x}yz =: \tilde{d}_t^3 , \quad \partial_{ty}^2 =: \tilde{\partial}^2_{zx} , \quad d\tilde{t} =: -\tilde{d}_{xyz} , \quad d^4 \tilde{t}xyz = \tilde{d}_1^4 . \quad (19)$$

This notation is convenient not only because it is shorter, but also because the twisted orientation is often physically very important. A twisted 3-covector $a d^3 \tilde{t}zy$ with $a > 0$ can for example represent the areic current of positive electric charge, or charge/(time \times area), across a yz plane. It is important to know whether this current is directed towards positive or negative x values, but this direction is not immediately evident from “ $\tilde{t}zy$ ”. The notation $a \tilde{d}_x^3$ makes this information immediately apparent instead, while still keeping the plane’s yz identity easily deducible.

As bases for the spaces of twisted vectors and covectors we take the twisted versions of (14) and (15). Note the following convenient relationships for multiplication by the 4-covector $d^4 \tilde{t}xyz \equiv \tilde{d}_1^4$ and 4-vector $\partial_{txyz}^4 \equiv \tilde{\partial}^4_1$:

$$\begin{aligned} \partial_t \cdot \tilde{d}_1^4 &= \tilde{d}_t^3 , & \partial_x \cdot \tilde{d}_1^4 &= \tilde{d}_x^3 , & \text{and so on,} \\ \partial_{tx} \cdot \tilde{d}_1^4 &= \tilde{d}_{tx}^3 , & \partial_{ty} \cdot \tilde{d}_1^4 &= \tilde{d}_{ty}^3 , & \text{and so on;} \\ \tilde{\partial}^4_1 \cdot \tilde{d}_t^3 &= \partial_t , & \tilde{\partial}^4_1 \cdot \tilde{d}_x^3 &= \partial_x , & \text{and so on,} \\ \tilde{\partial}^4_1 \cdot \tilde{d}_{tx}^3 &= \partial_{tx} , & \tilde{\partial}^4_1 \cdot \tilde{d}_{ty}^3 &= \partial_{ty} , & \text{and so on.} \end{aligned} \quad (20)$$

Bibliography

(“de X ” is listed under D , “van X ” under V , and so on, regardless of national conventions.)

Bossavit, A. (1991): *Differential Geometry: for the student of numerical methods in electromagnetism*. https://www.researchgate.net/publication/200018385_Differential_Geometry_for_the_student_of_numerical_methods_in_Electromagnetism.

Burke, W. L. (1983): *Manifestly parity invariant electromagnetic theory and twisted tensors*. J. Math. Phys. **24**¹, 65–69. [DOI:10.1063/1.525603](https://doi.org/10.1063/1.525603).

- (1987): *Applied Differential Geometry*, repr. (Cambridge University Press, Cambridge). doi:10.1017/CB09781139171786. First publ. 1985.
- (1995): *Div, Grad, Curl Are Dead*. http://people.ucsc.edu/~rmont/papers/Burke_DivGradCurl.pdf. See also <http://www.ucolick.org/~burke>.
- Davis, T. M., Lineweaver, C. H. (2004): *Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe*. Publ. Astron. Soc. Aust. **21**¹, 97–109. doi:10.1071/AS03040.
- Davis, T. M., Lineweaver, C. H., Webb, J. K. (2003): *Solutions to the tethered galaxy problem in an expanding universe and the observation of receding blueshifted objects*. Am. J. Phys. **71**⁴, 358–364. doi:10.1119/1.1528916.
- Eckart, C. (1940): *The thermodynamics of irreversible processes. III. Relativistic theory of the simple fluid*. Phys. Rev. **58**¹⁰, 919–924. doi:10.1103/PhysRev.58.919.
- Grøn, Ø. (2004): *Space geometry in rotating reference frames: a historical appraisal*. In: Rizzi, Ruggiero (2004): ch. 15:285–333. doi:10.1007/978-94-017-0528-8_17.
- Rizzi, G., Ruggiero, M. L., eds. (2004): *Relativity in Rotating Frames: Relativistic Physics in Rotating Reference Frames*. (Springer, Dordrecht). doi:10.1007/978-94-017-0528-8.
- Schouten, J. A. (1989): *Tensor Analysis for Physicists*, corr. 2nd ed. (Dover, New York). First publ. 1951.
- Weber, T. A. (2004): *Elementary considerations of the time and geometry of rotating reference frames*. In: Rizzi, Ruggiero (2004): ch. 7:139–153. doi:10.1007/978-94-017-0528-8_9.
- Whittaker, E. T. (1951): *A History of the Theories of Aether and Electricity*. Vol. 1: *The Classical Theories*, rev. and enl. ed. (Thomas Nelson and Sons, London). First publ. 1910.
- (1953): *A History of the Theories of Aether and Electricity*. Vol. 2: *The Modern Theories, 1900–1926*. (Thomas Nelson and Sons, London).