1 Lorentz and Poincaré Invariance

1.1 Introduction

The principles of special and general relativity are cornerstones of our understanding of the universe. The combination of gravity with quantum field theory remains an elusive goal and is outside the scope of our discussions. It will be sufficient for us to restrict our attention to special relativity, which is characterized by the Poincaré transformations. These are made up of the Lorentz transformations and spacetime translations. The Lorentz transformations are sometimes referred to as the homogeneous Lorentz transformations, since they preserve the origin, with the Poincaré transformations referred to as the inhomogeneous Lorentz transformations.

We classify fundamental particles by the representation of the Poincaré group that they transform under, which is related to their mass and intrinsic spin, e.g., spin-zero bosons, spin-half fermions and spin-one gauge bosons. Particles with integer spin (0, 1, 2, ...) are *bosons* and obey Bose-Einstein statistics and particles with half-integer spin (1/2, 3/2, 5/2, ...) are *fermions* and obey Fermi-Dirac statistics.

We first explore the nonrelativistic Galilean relativity that we are intuitively familiar with before we generalize to the case of special relativity.

1.1.1 Inertial Reference Frames

A *reference frame* is a three-dimensional spatial coordinate system with a clock at every spatial point. A *spacetime event* E is a point in spacetime,

$$E = (t, \mathbf{x}) = (t, x, y, z) = (t, x^1, x^2, x^3). \tag{1.1.1}$$

Note that we write the indices of the spatial coordinates $\mathbf{x} = (x^1, x^2, x^3)$ in the "up" position when discussing special relativity. The *trajectory* of a particle is then described in this reference frame by a continuous series of spacetime events; i.e., if $E = (t, \mathbf{x})$ is an element of the particle's trajectory, then the particle was at the point $\mathbf{x}$ at time t. When we discuss special relativity we will describe an event as $E = (ct, \mathbf{x})$, where c is the speed of light for reasons that will later become obvious. We refer to the set of all spacetime events as *Minkowski space*, $\mathbb{R}^{(1,3)}$, to indicate that there is one time dimension and three spatial dimensions.

When a particle or object is not subject to any external forces, we refer to it as *free*. An *inertial reference frame* is a reference frame in which a free particle or object moves in straight lines at a constant speed. An *inertial observer* is an observer who describes events in spacetime with respect to his chosen inertial reference frame. So an inertial observer describes the trajectory of a *free particle* as a straight line,

$$\mathbf{x}(t) = \mathbf{x}(0) + \mathbf{u}t, \tag{1.1.2}$$

where $\mathbf{u}$ is a constant three-vector, i.e., the constant velocity $\mathbf{u} = d\mathbf{x}/dt$.

Recall Newton's first law: "When viewed from any inertial reference frame, an object at rest remains at rest and an object moving continues to move with a constant velocity, unless acted upon by an external unbalanced force." So Newton's first law simply provides a definition of an inertial reference frame. It is in the inertial frames defined by the first law that Newton's second and third laws are formulated. In other words, Newton's laws have the same form in all inertial frames. It was assumed that all of the laws of physics have the same form in every inertial frame.

In arriving at his laws of motion Newton *assumed* that the displacement in time $\Delta t = t_2 - t_1$ between two events and all lengths would be the same in all frames. This is equivalent to Newton having assumed the laws of Galilean relativity.

1.1.2 Galilean Relativity

Consider any two observers $\mathcal{O}$ and $\mathcal{O}'$ with different reference frames such that a spacetime event E is described by observer $\mathcal{O}$ as $(t, \mathbf{x})$ and is described by observer $\mathcal{O}'$ as $(t', \mathbf{x}')$. These reference frames are not necessarily inertial.

If for every two spacetime events observers $\mathcal{O}$ and $\mathcal{O}'$ measure the same time difference and if they measure all lengths equal, then this means that

$$(t_2' - t_1') = (t_2 - t_1) \quad \text{and} \quad |\mathbf{x}_2'(t') - \mathbf{x}_1'(t')| = |\mathbf{x}_2(t) - \mathbf{x}_1(t)|, \tag{1.1.3}$$

where length is defined as the distance between two simultaneous events such as the locations of the end of a ruler at any given time. It then follows that the reference frames of $\mathcal{O}$ and $\mathcal{O}'$ must be related to each other by the transformation

$$t' = t + d \quad \text{and} \quad \mathbf{x}'(t') = O(t)\mathbf{x}(t) + \mathbf{b}(t), \tag{1.1.4}$$

where $O(t)$ is an orthogonal matrix at all times, $O^T(t) = O^{-1}(t)$.

Proof: Since $(t_2' - t_1') = (t_2 - t_1)$ for every two events we must have $t' = t + d$ where d is the difference in the settings of the clocks of the two observers.

Since the vectors $[\mathbf{x}_2'(t') - \mathbf{x}_1'(t')]$ and $[\mathbf{x}_2(t) - \mathbf{x}_1(t)]$ have the same length at all times, then they must be related by a rotation at all times; i.e., there must exist a matrix $O(t)$ that is orthogonal at all times so that

$$[\mathbf{x}_2'(t') - \mathbf{x}_1'(t')] = O(t)\,[\mathbf{x}_2(t) - \mathbf{x}_1(t)]\,. \tag{1.1.5}$$

In general there can be an arbitrary spatially independent function $\mathbf{b}(t)$ such that $\mathbf{x}'(t') = O(t)\mathbf{x}(t) + \mathbf{b}(t)$. The result is then proved.

If $\mathcal{O}$ and $\mathcal{O}'$ are inertial observers and if they measure the same time differences and lengths, then their reference frames are related by

$$t' = t + d \quad \text{and} \quad \mathbf{x}'(t') = O\mathbf{x}(t) - \boldsymbol{v}t + \mathbf{a}, \tag{1.1.6}$$

which is referred to as a *Galilean transformation*. Here d is the offset of the clock (i.e., *time translation*) of observer $\mathcal{O}'$ with respect to that of observer $\mathcal{O}$; the vector $(-\mathbf{a})$ is the displacement (i.e., *spatial translation*) of the origin of $\mathcal{O}'$ with respect to that of $\mathcal{O}$ at $t = 0$; $\boldsymbol{v}$ is the velocity (i.e., *velocity boost*) of $\mathcal{O}'$ with respect to $\mathcal{O}$; and O is the orthogonal matrix $(O^T = O^{-1})$ (i.e., *rotation*) that rotates the spatial axes of $\mathcal{O}'$ into the orientation of the spatial axes of $\mathcal{O}$.

Proof: If both observers are inertial, then they will each describe the trajectory of a free particle by a straight line

$$\mathbf{x}'(t') = \mathbf{x}'_0 + \mathbf{u}'t' \quad \text{and} \quad \mathbf{x}(t) = \mathbf{x}_0 + \mathbf{u}t, \tag{1.1.7}$$

where $\mathbf{x}$, $\mathbf{x}_0$, $\mathbf{u}$ and $\mathbf{u}'$ are all time-independent. Since they measure the same time differences and lengths, then from the above result we also have $t' = t + d$ and $\mathbf{x}'(t') = O(t)\mathbf{x}(t) + \mathbf{b}(t)$. It then follows that

$$\mathbf{x}'(t+d) = \mathbf{x}'_0 + \mathbf{u}'(t+d) = O(t)\mathbf{x}(t) + \mathbf{b}(t) = O(t)\left(\mathbf{x}_0 + \mathbf{u}t\right) + \mathbf{b}(t). \tag{1.1.8}$$

Acting with d/dt and using the notation $\dot{f} \equiv df/dt$ we find

$$\mathbf{u}' = \dot{O}(t)\left(\mathbf{x}_0 + \mathbf{u}t\right) + O(t)\mathbf{u} + \dot{\mathbf{b}}(t) \tag{1.1.9}$$

and acting with d/dt again gives

$$0 = \ddot{O}(t)\left(\mathbf{x}_0 + \mathbf{u}t\right) + 2\dot{O}(t)\mathbf{u} + \ddot{\mathbf{b}}(t). \tag{1.1.10}$$

This must be true for every free particle trajectory and hence must be true for every $\mathbf{u}$ and every $\mathbf{x}_0$, which means that we must have for every time t that

$$\ddot{O}(t) = \dot{O}(t) = 0 \quad \text{and so} \quad \ddot{\mathbf{b}}(t) = 0. \tag{1.1.11}$$

Then $O(t) = O$ is a time-independent orthogonal matrix and we can write $\mathbf{b}(t) = -\boldsymbol{v}t + \mathbf{a}$ for some time-independent $\boldsymbol{v}$ and $\mathbf{a}$, which proves Eq. (1.1.6).

At $t = 0$ the origin ($\mathbf{x}' = 0$) of the inertial reference frame of observer $\mathcal{O}'$ is at $\mathbf{x} = -\mathbf{a}$ in the inertial reference frame of $\mathcal{O}$ or equivalently the origin of $\mathcal{O}$ is at $\mathbf{x}' = \mathbf{a}$ in the frame of $\mathcal{O}'$. At t the origin of $\mathcal{O}'$ is at $\mathbf{x} = \boldsymbol{v}t - \mathbf{a}$, which shows that $\mathcal{O}'$ moves with a constant velocity $\boldsymbol{v}$ with respect to $\mathcal{O}$. Finally, O is the orthogonal matrix that rotates the spatial axes of $\mathcal{O}'$ to that of $\mathcal{O}$.

It was Newton's *assumption* that inertial frames are related by Galilean transformations and that all of the laws of physics are invariant under these transformations. Newton was assuming that the laws of physics were *Galilean invariant*.

An immediate consequence of Galilean invariance is the *Galilean law for the addition of velocities*, which is obtained by acting with d/dt on Eq. (1.1.6) to give

$$\mathbf{u}'(t') = O\mathbf{u}(t) - \boldsymbol{v}. \tag{1.1.12}$$

If we choose the two sets of axes to have the same orientation ($O = I$), then we recover the familiar Galilean addition of velocities $\mathbf{u}'(t) = \mathbf{u}(t) - \boldsymbol{v}$.

Einstein understood that there was an inconsistency between the assumed Galilean invariance of Newtonian mechanics and the properties of Maxwell's equations of electromagnetism. For example, if Galilean invariance applied to the propagation of light, then the velocity of light would depend on the boost $\boldsymbol{v}$; i.e., we would have $\mathbf{c}' = \mathbf{c} - \boldsymbol{v}$. In the nineteenth century it was thought that space was filled with some *luminiferous ether*[1] through which light propagated. The Earth's velocity through the ether changes as it rotates around the sun and hence light was expected to have a velocity dependent on direction. The Michelson-Morley experiment carried out in 1887 was designed to detect variations

[1] The term "luminiferous" means light-bearing and the original spelling of "aether" has widely been replaced by the modern English spelling "ether."

in the speed of light in perpendicular directions; however, it and subsequent experiments found no evidence of this.

1.2 Lorentz and Poincaré Transformations

Having understood Galilean relativity we can now generalize to the construction of the theory of special relativity. We first construct the Lorentz transformations and then build the Lorentz group and then extend this to the Poincaré group.

1.2.1 Postulates of Special Relativity and Their Implications

The postulates of special relativity are:

(i) The laws of physics are the same in all inertial frames.
(ii) The speed of light is constant and is the same in all inertial frames.

Let E_1 and E_2 be any two spacetime events labeled by inertial observer $\mathcal{O}$ as $x_1^\mu = (ct_1, \mathbf{x}_1)$ and $x_2^\mu = (ct_2, \mathbf{x}_2)$, respectively. The *spacetime displacement* of these two events in the frame of $\mathcal{O}$ is defined as

$$z^\mu \equiv \Delta x^\mu = (x_2^\mu - x_1^\mu) = (c(t_2 - t_1), \mathbf{x}_2 - \mathbf{x}_1) = (c\Delta t, \Delta\mathbf{x}). \tag{1.2.1}$$

Consider the special case where event E_1 consists of a flash of light and event E_2 refers to the arrival of that flash of light. Since c is the speed of light, then $c = \Delta\mathbf{x}/\Delta t$ and so $(z^0)^2 - \mathbf{z}^2 = (c\Delta t)^2 - (\Delta\mathbf{x})^2 = 0$. Let $\mathcal{O}'$ be any other inertial observer who in his frame records these same two events as $x_1'^\mu = (ct_1', \mathbf{x}_1')$ and $x_2'^\mu = (ct_2', \mathbf{x}_2')$, respectively. From postulate (ii) it follows that $c = \Delta\mathbf{x}'/\Delta t'$ and hence that $z'^2 = (c\Delta t')^2 - (\Delta\mathbf{x}')^2 = 0$.

The Minkowski-space *metric tensor*, $g_{\mu\nu}$, is defined[2] by

$$g_{00} = -g_{11} = -g_{22} = -g_{33} = +1 \qquad \text{and} \qquad g_{\mu\nu} = 0 \text{ for } \mu \neq \nu, \tag{1.2.2}$$

which we can also write as $g_{\mu\nu} \equiv \mathrm{diag}(1, -1, -1, -1)$. We can then further define

$$x^2 \equiv x^\mu g_{\mu\nu} x^\nu = (x^0)^2 - \mathbf{x}^2 = (ct)^2 - \mathbf{x}^2, \tag{1.2.3}$$

where we use the *Einstein summation convention* that repeated spacetime indices are understood to be summed over. In general relativity spacetime is curved and the metric tensor becomes spacetime-dependent, $g_{\mu\nu} \to g(x)_{\mu\nu}$.

For our special case consisting of the emission and reception of light at events E_1 and E_2, respectively, we say that the events have a *lightlike* displacement. Using the above notation we see that the spacetime interval in this case satisfies the condition that $z'^2 = z^2 = 0$, where $z^2 = (\Delta x)^2 = (x_2 - x_1)^2$ and $z'^2 = (\Delta x')^2 = (x_2' - x_1')^2$. If two events have a lightlike displacement in one inertial frame, then by the postulates of relativity they have a lightlike displacement in every frame. The set of all lightlike displacements makes up something that is referred to as the *light cone*.

[2] In general relativity and some quantum field theory textbooks $g_{\mu\nu} \equiv \mathrm{diag}(-1, +1, +1, +1)$ is used. The more common particle physics choice is the metric tensor defined here.

Cambridge University Press & Assessment
978-1-108-47090-2 — Introduction to Quantum Field Theory
Anthony G. Williams
Excerpt
More Information

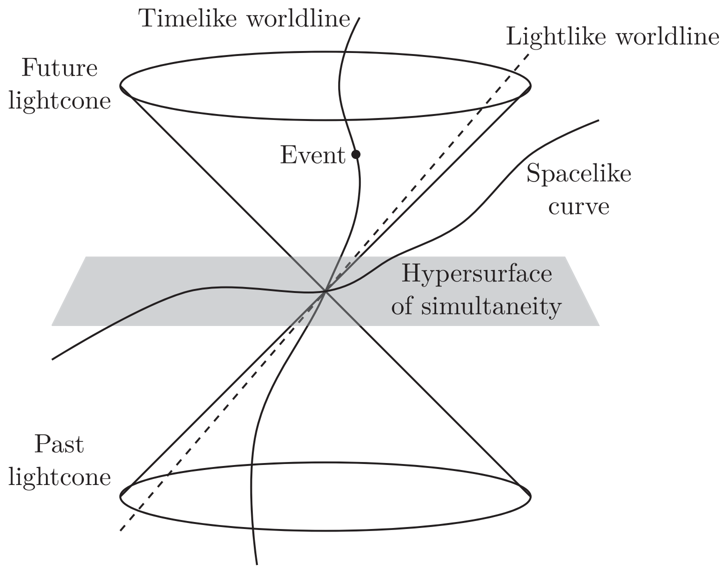

Figure 1.1 The light cone at a point on a particle worldline. In order to graphically represent the light cone, we suppress one of the three spatial dimensions.

The *spacetime interval* between spacetime events E_1 and E_2 is defined as

$$s^2 \equiv z^2 = (\Delta x)^2 = (z^0)^2 - \mathbf{z}^2 = (c\Delta t)^2 - (\Delta \mathbf{x})^2 = c^2(t_2 - t_1)^2 - (\mathbf{x}_2 - \mathbf{x}_1)^2. \tag{1.2.4}$$

If $s^2 = 0$ the events have a lightlike displacement and they lie on the light cone.

To visualize the light cone and its usefulness in categorizing the spacetime interval, it is useful to temporarily suppress one of the spatial dimension as in Fig. 1.1. It is easy to see that all events E_2 with a lightlike displacement from event E_1 will map out a cone in the three-dimensional spacetime with the time axis as the central axis of the future and past light cones and with E_1 at the point where the light cones touch. The trajectory of a particle is a continuous sequence of spacetime events, which in special relativity is often referred to as the *worldline* of the particle.

If a particle contains event E_1, then any subsequent event E_2 on its worldline must lie inside the forward light cone of E_1. Similarly, in such a case if E_2 occurred on the worldline before E_1 then E_2 must lie inside the backward light cone. Being inside the light cone means that $s^2 > 0$ and we describe such events as *timelike* separated events. If $s^2 < 0$, then we say that the two events have a *spacelike* displacement and no particle moving at or below the speed of light could be present at both events.

We will later see that there is always an inertial reference frame where spacelike separated events are simultaneous. Simultaneous events cannot have a *causal* relationship. In addition, postulate (i) tells us that causal relationships should not depend on the observer's inertial reference frame. Thus spacelike separated events cannot be causally related; i.e., if E_1 and E_2 have a spacelike displacement, then what happens at E_1 cannot influence what happens at E_2 and vice versa.

A particle cannot be in two places at the same time in any inertial frame and so no part of its worldline can be spacelike; i.e., particles cannot travel faster than the speed of light. We will see in Sec. 2.6 that it would take an infinite amount of energy to accelerate a massive particle to the speed of light and so massive particles have worldlines that must lie entirely *within* the light cone; i.e., they have timelike worldlines. We will also see that massless particles must travel at the velocity of light and so their worldlines, like that of light, must lie entirely *on* the light cone.

Consider the special case where at $t = 0$ the clocks of the two observers are synchronized and the origins and axes of their coordinate systems coincide; i.e., at $t = 0$ the spacetime origins of $\mathcal{O}$ and $\mathcal{O}'$ are coincident, $x^\mu = x'^\mu = 0$. It must be possible to define an invertible (one-to-one and onto) transformation between their coordinate systems of the form $x^\mu \to x'^\mu = x'^\mu(x)$, where

$$dx^\mu \to dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu \equiv \Lambda^\mu{}_\nu dx^\nu \quad \text{or in matrix form} \quad dx \to dx' = \Lambda\, dx, \tag{1.2.5}$$

where we have defined $\Lambda^\mu{}_\nu \equiv \partial x'^\mu / \partial x^\nu$. The sixteen elements of Λ must be real; i.e., $\Lambda^\mu{}_\nu \in \mathbb{R}$ for all μ and $\nu = 0, 1, 2, 3$. In the matrix notation that we sometimes use x^μ is the element in row μ of a four-component column vector x, and $\Lambda^\mu{}_\nu$ is the element in row μ and column ν of the 4×4 real Lorentz transformation matrix Λ; i.e., column vector x has elements x^{row} and the matrix Λ has elements $\Lambda^{\text{row}}{}_{\text{column}}$. Note that for historical reasons we refer to x^μ as a *contravariant* four-vector. The matrices Λ with elements $\Lambda^\mu{}_\nu$ are referred to as *Lorentz transformations*.

An important and useful concept is that of the *proper time*[3] for a moving object, which is the time showing on a clock that is carried along with the object. Also important is the concept of the *proper length* of an object that is its length in its own rest frame. For a free particle we define the *comoving inertial frame* as the inertial frame in which the particle remains at rest. For an object that is accelerating we can still define an *instantaneously comoving inertial frame* as an inertial frame in which the accelerating particle is at rest at a particular instant of time.

The discussion so far allows for the possibility that the transformations $\Lambda^\mu{}_\nu$ might depend on x^μ; i.e., we might have $\Lambda^\mu{}_\nu(x)$. However from postulate (i) it follows that the Lorentz transformations $\Lambda^\mu{}_\nu$ must be spacetime independent; i.e., *any Lorentz transformation matrix* Λ *contains sixteen constants.* So for any x^μ we have

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu \qquad \text{or in matrix form} \qquad x \to x' = \Lambda x. \tag{1.2.6}$$

Proof: Consider a free particle being observed by two inertial observers $\mathcal{O}$ and $\mathcal{O}'$. Both observers will describe the worldline as a straight line. Since by postulate (i) Minkowski space must be homogeneous, then an inertial observer must see equal increments of dx^μ in equal increments of the free particle's proper time $d\tau$; i.e., for a free particle we must have that

$$\frac{dx^\mu}{d\tau} \quad \text{and} \quad \frac{dx'^\mu}{d\tau} \quad \text{are constants} \quad \Rightarrow \quad \frac{d^2x^\mu}{d\tau^2} = \frac{d^2x'^\mu}{d\tau^2} = 0. \tag{1.2.7}$$

We now consider

$$\frac{dx'^\mu}{d\tau} = \frac{\partial x'^\mu}{\partial x^\nu}\frac{dx^\nu}{d\tau} = \Lambda^\mu{}_\nu \frac{dx^\nu}{d\tau} \tag{1.2.8}$$

and differentiate with respect to the proper time τ to give

$$\frac{d^2x'^\mu}{d\tau^2} = \frac{\partial x'^\mu}{\partial x^\nu}\frac{d^2x^\nu}{d\tau^2} + \frac{\partial^2 x'^\mu}{\partial x^\rho \partial x^\nu}\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau}. \tag{1.2.9}$$

[3] The meaning of the English word *proper* in this context originates from the French word *propre* meaning *own* or *one's own*. The proper time is the time belonging to the object itself.

Cambridge University Press & Assessment
978-1-108-47090-2 — Introduction to Quantum Field Theory
Anthony G. Williams
Excerpt
More Information

This must be true for all $dx^\mu/d\tau$. Then from Eq. (1.2.7) we immediately find

$$\frac{\partial^2 x'^\mu}{\partial x^\rho \partial x^\nu} = \frac{\partial}{\partial x^\rho}\Lambda^\mu{}_\nu = 0 \tag{1.2.10}$$

and so $\Lambda^\mu{}_\nu = \partial x'^\mu/\partial x^\nu$ is independent of x; i.e., Λ is a matrix of constants. Then dividing x^μ into an infinite number of infinitesimals gives $x'^\mu = \Lambda^\mu{}_\nu x^\nu$.

It follows that the most general transformation between two inertial frames is

$$x^\mu \to x'^\mu = p(x) \equiv \Lambda^\mu{}_\nu x^\nu + a^\mu \quad \text{or in matrix form} \quad x \to x' = \Lambda x + a, \tag{1.2.11}$$

which is referred to as a *Poincaré transformation*. In his frame observer $\mathcal{O}'$ describes the spacetime origin of $\mathcal{O}$ as being at a^μ. So a^μ is a *spacetime translation*.

Let us return to our example of the two events linked by the propagation of light as seen in two different frames; i.e., they have a lightlike displacement. Then from postulate (ii) and the above arguments we have $z^2 = z'^2 = 0$ and so

$$z'^2 = z'^T g z' = z^T \Lambda^T g \Lambda z = z^2 = z^T g z = 0, \tag{1.2.12}$$

where $z^\mu = (z, \mathbf{z})$ with $z \equiv |\mathbf{z}|$. Then noting that $\Lambda^T g \Lambda$ is a real symmetric matrix and using the theorem given below, we see that for every Lorentz transformation Λ we must have $\Lambda^T g \Lambda = a(\Lambda) g$ for some real constant $a(\Lambda)$ independent of $\mathbf{z}$. For $\Lambda = I$ we have $I^T g I = g$ and so $a(I) = 1$. Then we have $1 = a(I) = a(\Lambda^{-1}\Lambda) = a(\Lambda^{-1})a(\Lambda)$ and so $a(\Lambda^{-1}) = 1/a(\Lambda)$. If observer $\mathcal{O}'$ is related to observer $\mathcal{O}$ by Λ, then $\mathcal{O}$ is related to $\mathcal{O}'$ by Λ^{-1}. The labeling of inertial observers is physically irrelevant, since there is no preferred inertial frame. So $a(\Lambda) = a(\Lambda^{-1}) = 1/a(\Lambda)$ and $a(\Lambda) = \pm 1$ for all Λ. Since $a(I) = 1$ and since $a(\Lambda'\Lambda) = a(\Lambda')a(\Lambda) = a(\Lambda')/a(\Lambda) = a(\Lambda'\Lambda^{-1})$ for all Λ' and Λ, then $a(\Lambda) = 1$ for all Λ. So *any* Λ must satisfy

$$\Lambda^T g \Lambda = g \qquad \text{or equivalently} \qquad \Lambda^\sigma{}_\mu g_{\sigma\tau} \Lambda^\tau{}_\nu = g_{\mu\nu}. \tag{1.2.13}$$

This is equivalent to the statement that *the Minkowski-space metric tensor $g_{\mu\nu}$ is invariant under Lorentz transformations*.

Theorem: Let A be a real symmetric matrix, independent of $\mathbf{z}$ and define $z \equiv |\mathbf{z}|$, then if

$$\begin{pmatrix} z & \mathbf{z}^T \end{pmatrix} A \begin{pmatrix} z \\ \mathbf{z} \end{pmatrix} = 0 \qquad \text{for all} \quad \mathbf{z} \tag{1.2.14}$$

there must exist some real constant a independent of $\mathbf{z}$ such that $A = ag$, where g is the Minkowski-space metric tensor, $g = \mathrm{diag}(1, -1, -1, -1)$.
Proof: Since A is a real symmetric matrix, then we can write it as

$$A \equiv \begin{pmatrix} a & \mathbf{b}^T \\ \mathbf{b} & C \end{pmatrix}, \tag{1.2.15}$$

where $a \in \mathbb{R}$, $\mathbf{b}$ is a real three-vector and C is a real symmetric 3×3 matrix. We then obtain $az^2 + 2z\mathbf{z} \cdot \mathbf{b} + \mathbf{z}^T C \mathbf{z} = 0$, which must be true for all $\mathbf{z}$. The only way this can be true for both $\pm\mathbf{z}$ is if $\mathbf{b} = 0$. Then we have $\mathbf{z}^T (aI + C)\, \mathbf{z} = 0$ for all $\mathbf{z}$ and since $(aI + C)$ is a real symmetric matrix, this means that we must have $C = -aI$. This then proves that $A = ag$ as required.

Since $\det(g) = \det(\Lambda^T g \Lambda) = \det(\Lambda)^2 \det(g)$ and Λ is a real matrix, then

$$\det(\Lambda) = \pm 1. \tag{1.2.16}$$

We *define* $g^{\mu\nu}$ to be the matrix elements of the matrix g^{-1}, i.e.,

$$g^{-1} g = I \qquad \text{or equivalently} \qquad g^{\mu\sigma} g_{\sigma\nu} \equiv \delta^\mu{}_\nu, \tag{1.2.17}$$

where the $\delta^\mu{}_\nu$ are the elements of the identity matrix, i.e., $\delta^\mu{}_\nu = 1$ if $\mu = \nu$ and zero otherwise. This is also what is done in general relativity. For the Minkowski-space metric $g = \mathrm{diag}(1, -1, -1, -1)$ that is relevant for special relativity it is clear that $g^{-1} = g$, since $g^2 = I$. So $g^{\mu\nu}$ and $g_{\mu\nu}$ are numerically equal, i.e.,

$$g^{-1} = g \qquad \text{or equivalently} \qquad g^{\mu\nu} = g_{\mu\nu}. \tag{1.2.18}$$

Acting on Eq. (1.2.13) from the left with $g^{\rho\mu}$ we find that

$$g\Lambda^T g \Lambda = gg = I \qquad \text{or equivalently} \qquad g^{\rho\mu} \Lambda^\sigma{}_\mu g_{\sigma\tau} \Lambda^\tau{}_\nu = g^{\rho\mu} g_{\mu\nu} = \delta^\rho{}_\nu. \tag{1.2.19}$$

Note that since $\det \Lambda \neq 0$ then Λ^{-1} exists. Since $g\Lambda^T g \Lambda = I$ then acting from the left with Λ and from the right with Λ^{-1} gives $\Lambda(g\Lambda^T g) = I$. Then $\Lambda(g\Lambda^T g) = (g\Lambda^T g)\Lambda = I$ for all Λ and so $(g\Lambda^T g)$ is the inverse of Λ,

$$\Lambda^{-1} = g\Lambda^T g \qquad \text{or equivalently} \qquad (\Lambda^{-1})^\rho{}_\lambda = g^{\rho\mu} \Lambda^\sigma{}_\mu g_{\sigma\lambda}. \tag{1.2.20}$$

Furthermore, note that since $\Lambda g \Lambda^T g = I$ and $g^2 = I$ then $\Lambda g \Lambda^T = g$ and so comparing with Eq. (1.2.13) we see that if Λ is a Lorentz transform, then so is Λ^T.

The x^μ are the contravariant four-vector spacetime coordinates. Any contravariant four-vector transforms by definition as $V^\mu \to V'^\mu = \Lambda^\mu{}_\nu V^\nu$. A *covariant* four-vector by definition transforms with the inverse transformation, i.e., $V_\mu \to V'_\mu = V_\nu (\Lambda^{-1})^\nu{}_\mu$ so that $V_\mu V^\mu = V^\mu V_\mu$ is invariant under the transformation, i.e.,

$$V'_\mu V'^\mu = V_\nu (\Lambda^{-1})^\nu{}_\mu \Lambda^\mu{}_\sigma V^\sigma = V_\nu \delta^\nu{}_\sigma V^\sigma = V_\nu V^\nu. \tag{1.2.21}$$

An arbitrary *tensor* with m contravariant and n covariant indices transforms as

$$\begin{aligned} &T^{\mu_1\mu_2\ldots\mu_m}{}_{\nu_1\nu_2\ldots\nu_n} \to (T')^{\mu_1\mu_2\ldots\mu_m}{}_{\nu_1\nu_2\ldots\nu_n} \\ &\quad = \Lambda^{\mu_1}{}_{\sigma_1} \Lambda^{\mu_2}{}_{\sigma_2} \cdots \Lambda^{\mu_m}{}_{\sigma_m} T^{\sigma_1\sigma_2\ldots\sigma_m}{}_{\tau_1\tau_2\ldots\tau_n} (\Lambda^{-1})^{\tau_1}{}_{\nu_1} (\Lambda^{-1})^{\tau_2}{}_{\nu_2} \cdots (\Lambda^{-1})^{\tau_n}{}_{\nu_n} \end{aligned} \tag{1.2.22}$$

and we refer to such a tensor as having rank $(m+n)$ and being of type (m,n). Note that if we *contract* a contravariant and covariant index by setting them equal and summing over them, then we obtain

$$\begin{aligned} &T^{\mu_1\ldots\mu_{i-1}\rho\mu_{i+1}\ldots\mu_m}{}_{\nu_1\ldots\nu_{j-1}\rho\nu_{j+1}\ldots\nu_n} \to (T')^{\mu_1\ldots\mu_{i-1}\rho\mu_{i+1}\ldots\mu_m}{}_{\nu_1\ldots\nu_{j-1}\rho\nu_{j+1}\ldots\nu_n} \\ &\quad = \Lambda^{\mu_1}{}_{\sigma_1} \cdots \Lambda^{\rho}{}_{\sigma_i} \cdots \Lambda^{\mu_m}{}_{\sigma_m} T^{\sigma_1\sigma_2\ldots\sigma_m}{}_{\tau_1\tau_2\ldots\tau_n} (\Lambda^{-1})^{\tau_1}{}_{\nu_1} \cdots (\Lambda^{-1})^{\tau_j}{}_{\rho} \cdots (\Lambda^{-1})^{\tau_n}{}_{\nu_n} \\ &\quad = \Lambda^{\mu_1}{}_{\sigma_1} \cdots \Lambda^{\mu_{i-1}}{}_{\sigma_{i-1}} \Lambda^{\mu_{i+1}}{}_{\sigma_{i+1}} \cdots \Lambda^{\mu_m}{}_{\sigma_m} T^{\sigma_1\ldots\sigma_{i-1}\lambda\sigma_{i+1}\ldots\sigma_m}{}_{\tau_1\ldots\tau_{j-1}\lambda\tau_{j+1}\ldots\tau_n} \\ &\qquad \times (\Lambda^{-1})^{\tau_1}{}_{\nu_1} \cdots (\Lambda^{-1})^{\tau_{j-1}}{}_{\nu_{j-1}} (\Lambda^{-1})^{\tau_{j+1}}{}_{\nu_{j+1}} \cdots (\Lambda^{-1})^{\tau_n}{}_{\nu_n}, \end{aligned} \tag{1.2.23}$$

where we have used $\Lambda^\rho{}_{\sigma_i} (\Lambda^{-1})^{\tau_j}{}_\rho = \delta^{\tau_j}{}_{\sigma_i}$. We see that if a type (m,n) tensor has a pair of spacetime indices contracted, then it becomes a type $(m-1, n-1)$ tensor.

We observe that $g_{\mu\nu}$ is invariant under Lorentz transformations, since

$$(\Lambda^{-1})^\sigma{}_\mu g_{\sigma\tau} (\Lambda^{-1})^\tau{}_\nu = g_{\mu\nu}. \tag{1.2.24}$$

This follows in any case since Eq. (1.2.13) is true for any Lorentz transformation.

We can use the properties of the Lorentz transformations to establish a simple relationship between the covariant and contravariant vectors. Let V^μ be a contravariant four-vector, then $(V^T g)$ is a row matrix with elements $V^\nu g_{\nu\mu}$. Under a Lorentz transformation we find that $(V^T g)$ transforms as

$$(V^T g) \to (V'^T g) = V^T \Lambda^T g = V^T g g \Lambda^T g = (V^T g)\Lambda^{-1} \tag{1.2.25}$$

or in index notation

$$\begin{aligned}(V^\nu g_{\nu\mu}) \to (V'^\nu g_{\nu\mu}) &= V^\rho \Lambda^\nu{}_\rho g_{\nu\mu} = V^\tau g_{\tau\sigma} g^{\sigma\rho} \Lambda^\nu{}_\rho g_{\nu\mu} = (V^\tau g_{\tau\sigma})(\Lambda^{-1})^\sigma{}_\mu, \\ \Rightarrow \quad (V^T g)V &= (V^\nu g_{\nu\mu})V^\mu = V^2.\end{aligned} \tag{1.2.26}$$

So for Lorentz transformations we can identify the elements of the covariant four-vector V_μ with the elements of the row vector $(gV)^T$, i.e.,

$$V_\mu \equiv (V^T g)_\mu = V^\nu g_{\nu\mu}. \tag{1.2.27}$$

Since $g^2 = I$ it then follows that

$$V^\mu = V_\nu g^{\nu\mu} \qquad \text{or equivalently} \qquad V = ((V^T g)g)^T, \tag{1.2.28}$$

where here the index notation is more concise than the matrix notation. These arguments extend to any spacetime indices on an arbitrary tensor. Hence we can use $g^{\mu\nu}$ to *raise* a covariant index to be a contravariant one and conversely $g_{\mu\nu}$ can be used to *lower* a contravariant index to be a covariant one, e.g.,

$$T^{\mu_1 \dots}{}_\mu{}^{\dots\mu_m}{}_{\nu_1\dots}{}^\nu{}_{\dots\nu_n} = g_{\mu\mu_i} g^{\nu\nu_j} T^{\mu_1\dots\mu_i\dots\mu_m}{}_{\nu_1\dots\nu_j\dots\nu_n}, \tag{1.2.29}$$

where every contravariant ("up") index transforms with a Λ and every covariant ("down") index transforms with a Λ^{-1}. If a type (m, m) tensor has all of its m pairs of indices contracted, then the result is *Lorentz invariant*; e.g., if V^μ and W^μ are four-vectors, then $V^\mu W_\nu$ is a $(1, 1)$ tensor and we have Lorentz invariants,

$$\begin{aligned}V \cdot W &\equiv V^\mu W_\mu = V_\mu W^\mu = V^\mu g_{\mu\nu} W^\nu = V_\mu g^{\mu\nu} W_\nu, \\ V^2 &\equiv V \cdot V = V^\mu V_\mu = V_\mu V^\mu = V^\mu g_{\mu\nu} V^\nu = V_\mu g^{\mu\nu} V_\nu.\end{aligned} \tag{1.2.30}$$

If the spacetime interval between any two events E_A and E_B is Lorentz invariant and translationally invariant, then it is Poincaré invariant. This follows since if $x' = \Lambda x + a$, then

$$s'^2 = z'^2 = (x'_B - x'_A)^2 = (\Lambda x_B - \Lambda x_A)^2 = (x_B - x_A)^2 = z^2 = s^2. \tag{1.2.31}$$

Note that *any* two spacetime displacements z'^μ and z^μ with the same interval, $z'^2 = z^2 = s^2$, are related by some Lorentz transformation by definition, $z' = \Lambda z$.

With z^0 on the y-axis and $|\mathbf{z}|$ on the x-axis constant s^2 corresponds to hyperbolic curves in (a) the forward light cone if $s^2 > 0$ and $z^0 > 0$, (b) the backward light cone if $s^2 > 0$ and $z^0 < 0$, and (c) the spacelike region if $s^2 < 0$.

$$\begin{aligned}&\text{If} \quad z^2 = s^2 = d > 0 \quad \text{then there is a} \quad z'^\mu = (\pm\sqrt{d}, \mathbf{0}) = (\pm\Delta\tau, \mathbf{0}) \\ &\text{if} \quad z^2 = s^2 = -d < 0 \quad \text{then there is a} \quad z'^\mu = (0, \sqrt{d}\mathbf{n}) = (0, L_0\mathbf{n}),\end{aligned} \tag{1.2.32}$$

where d is a positive real number, $\Delta\tau$ is the proper time between the events, $\mathbf{n}$ is an arbitrary unit vector, and L_0 is the proper length separating the events.

Since by definition all four-vectors transform the same way under Lorentz transformations, then we have the general results that (i) *for any timelike four-vector we can always find an inertial reference*

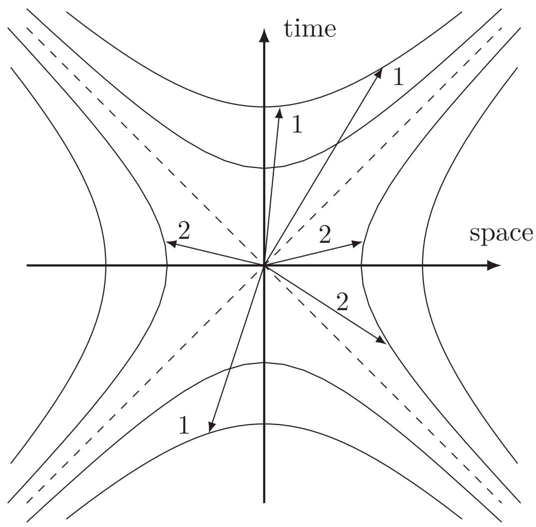

Figure 1.2 A two-dimensional representation of spacetime. Displacement vectors labeled by the same number (1 or 2) are related by Lorentz transformations.

frame where the spatial component vanishes and the time component is the proper time component, and (ii) *for any spacelike four-vector we can always find an inertial reference frame where the time component vanishes and the spatial component has its proper length* (see Fig. 1.2).

The *Lorentz transformations form a group* since they satisfy the four conditions that define a group: (i) *closure:* if Λ_1 and Λ_2 are two arbitrary Lorentz transformations, then $g = \Lambda_1^T g \Lambda_1 = \Lambda_2^T g \Lambda_2$ and hence $(\Lambda_1\Lambda_2)^T g(\Lambda_1\Lambda_2) = \Lambda_2^T \Lambda_1^T g \Lambda_1 \Lambda_2 = \Lambda_2^T g \Lambda_2 = g$ and hence $(\Lambda_1\Lambda_2)$ is also a Lorentz transformation; (ii) *associativity:* since consecutive Lorentz transformations are given by matrix multiplication, and since matrix multiplication is associative, then Lorentz transformations are also, $\Lambda_1(\Lambda_2\Lambda_3) = (\Lambda_1\Lambda_2)\Lambda_3$; (iii) *identity:* $\Lambda = I$ is a Lorentz transformation since $I^T g I = g$; and (iv) *inverse:* for every Lorentz transformation the group contains an inverse $\Lambda^{-1} = g\Lambda^T g$ as shown above. The Lorentz group is typically denoted as $O(1,3)$ and is a generalization of the rotation group $O(3)$ to four dimensions with a Minkowski-space metric. It is an example of a pseudo-orthogonal group.

Mathematical digression: Orthogonal and pseudo-orthogonal groups: Both orthogonal transformations and Lorentz transformations have the property that they are made up of real matrices A that satisfy $A^T G A = G$, where G is the relevant "metric" tensor. For the orthogonal transformations, $O(3)$, in their defining representation we have that A are real 3×3 matrices and G is the 3×3 identity matrix, i.e., with three entries of $+1$ down the diagonal such that $A^T A = I$. For the Lorentz transformations we have $G = g$, where g is given by Eq. (1.2.2) with a $+1$ and three -1 entries down the diagonal leading to the notation $O(1,3)$. The meaning of the group $O(m,n)$ in its defining representation is then straightforward to understand; it is the group of $(n+m) \times (n+m)$ real matrices A that satisfy $A^T G A = G$ with the diagonal matrix G having m entries of $+1$ and n entries of -1. If $n = 0$ then the group is the orthogonal group $O(m)$, otherwise for $n \neq 0$ we refer to the group $O(m,n)$ as *pseudo-orthogonal* or sometimes as *indefinite orthogonal*. Here we are using the "timelike" definition of the metric tensor g with $g^{00} = -g^{11} = -g^{22} = -g^{33} = +1$, which suggests writing the Lorentz transformations as