Spacetime and Geometry - Sean M. Carroll

Notes by Gene

2026

Contents

1 Special Relativity and Flat Spacetime (pp. 1 - 47)

1.1 Introductory Concepts

General relativity: basically gravity and geodesics.

- Most forces are embedded in spacetime (e.g. Electromagnetic field).
- "Gravity" isn't a force, but the *curvature* of spacetime. It's an *inherent* feature.

Note: We adopt the convention $c = 1$ for the speed of light.

We begin from Newton's law of gravitation: $F = GMm/r^2$. This can be rewritten as Poisson's Equation, describing gravity in terms of a potential.

$$\underbrace{\nabla^2\Phi}_{curvature} = \underbrace{4\pi G\rho}_{source}$$

- Φ is the gravitational potential, a scalar field of potential energy per mass.
- ρ is the mass density, i.e. the distribution of mass in space. This is the source that "generates" the gravitational potential.
- Acceleration is simple: $a = \nabla\Phi$, the gradient.

Roughly, mass affects the shape of the gravitational potential field. Similarly, in general relativity, spacetime curves in the presence of matter and energy.

However, Poisson's equation (and Newton's law) describes an instantaneous transfer of information: a change in mass immediately affects gravity. This can't be true, if special relativity is accepted. So the equations must be modified.

Later we'll derive

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}$$
$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0$$

The first equation is Einstein's equation. The second describes a geodesic, where $x^\mu(\lambda)$ is the parametrized path.

1.2 Special Relativity

1.2.1 Basics

(Largely a review) Special relativity is a theory of 4D spacetime, the “Minkowski Space” (a special case of 4D manifolds). The distinction between “space” and “time” is arbitrary — you have to think of “spacetime” as a whole entity, not the sum of its parts.

In Newtonian dynamics, we have a well-defined notion of “simultaneity.” In SR, we don’t. Particles stay within their light cones in a spacetime diagram.

In Cartesian coordinates, distance is invariant under coordinate changes:

$$(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2$$

In SR, the “spacetime interval” is invariant under coordinate changes:

$$(\Delta s)^2 = -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

We sum over Greek indices for spacetime coordinates and Latin indices for space coordinates. We also define the spacetime metric $\eta_{\mu\nu}$.

$$x^\mu : \begin{cases} x^0 = ct \\ x^1 = x \\ x^2 = y \\ x^3 = z \end{cases} \quad , \quad x^i : \begin{cases} x^1 = x \\ x^2 = y \\ x^3 = z \end{cases}$$

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Then we can rewrite the spacetime interval more compactly. The summation convention is that indices in both superscript & subscript are summed over.

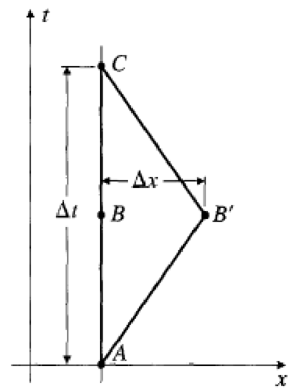
$$(\Delta s)^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu$$

$$\begin{aligned} (\Delta s)^2 &< 0 : \text{timelike separated} \\ (\Delta s)^2 &= 0 : \text{null} \\ (\Delta s)^2 &> 0 : \text{spacelike separated} \end{aligned}$$

Define “proper time” τ to measure the time elasped as seen by an observer moving in a straight path between two events. I.e. This is the time recorded by the *observer*. We add a minus sign because otherwise, timelike separated events would be negative and that’s weird.

$$(\Delta \tau)^2 = -(\Delta s)^2 = -\eta_{\mu\nu} \Delta x^\mu \Delta x^\nu$$

1.2.2 The Twin Paradox



The Twin Paradox

Let $B = \left(\frac{1}{2}\Delta t, 0\right)$ and $B' = \left(\frac{1}{2}\Delta t, \Delta x\right)$ with $\Delta x = \frac{1}{2}v\Delta t$.

Alice travels straight from A to B. Bob travels from A to B'. The proper times (with $c = 1$ and $v \leq c$) are

$$\text{Alice} : \Delta\tau_{AB} = \frac{1}{2}\Delta t$$

$$\text{Bob} : \Delta\tau_{AB'} = \sqrt{\left(\frac{1}{2}\Delta t\right)^2 - (\Delta x)^2} = \frac{1}{2}\Delta t\sqrt{1 - v^2}$$

Bob has aged less than Alice! Because a nonstraight path in *spacetime* has a different interval than a straight path. Additionally, the elasped time for the observer depends on the path you take.

1.2.3 Lorentz Transformations

Transformations in spacetime. The basic example is a translation with δ as the potential. We "shift" one variable, either spatial or temporal. Note that the distance Δx^μ between two points remains the same, thus the interval $(\Delta s)^2$ is invariant.

$$x^\mu \longrightarrow x^{\mu'} = \delta^{\mu'}_{\mu}\left(x^\mu + a^\mu\right), \quad \delta^{\mu'}_{\mu} = \begin{cases} 1 & \mu' = \mu \\ 0 & \mu' \neq \mu \end{cases}$$

Other examples are rotations and boosts (offsets by a constant velocity vector). We model these with a matrix Λ.

$$x^{\mu'} = \Lambda^\mu_\nu x^\nu, \quad \text{i.e.} \quad x' = \Lambda x$$

The interval $(\Delta s)^2$ must remain invariant, so Δx^μ must not change. Thus we have

$$\begin{gathered}(\Delta s)^2 = (\Delta x)^T \eta (\Delta x) = (\Delta x')^T \eta (\Delta x') \\ (\Delta x)^T \eta (\Delta x) = (\Delta x)^T \Lambda^T \eta \Lambda (\Delta x) \\ \eta = \Lambda^T \eta \Lambda\end{gathered}$$

The Λ that satisfies the above is a Lorentz transformation. Under matrix multiplication, the Lorentz transformations form a Lorentz group. The set of translations and Lorentz transformations form a 10-parameter, non-Abelian group known as the *Poincaré group*.

As an example, consider a boost in the x-direction (which can be thought of as rotations between space and time directions).

$$\Lambda^{\mu'}_{\nu} = \begin{pmatrix} \cosh\phi & -\sinh\phi & 0 & 0 \\ -\sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \phi \in (-\infty, \infty)$$

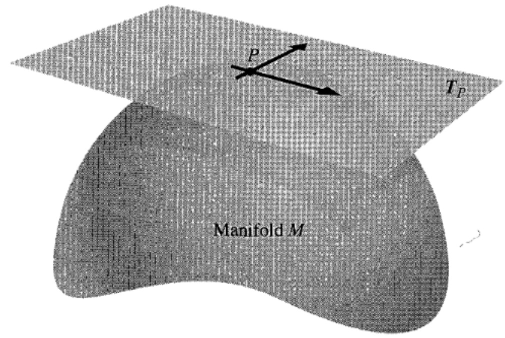
By applying Λ, we get $t' = t\cosh\phi - x\sinh\phi$ and $x' = -t\sinh\phi + x\cosh\phi$. Focusing on the point at x', we see that it moves according to $t\sinh\phi = x\cosh\phi$. Thus $v = x/t = \tanh\phi$.

Using $\phi = \tanh^{-1} v$, we can reconstruct the usual Lorentz transformations. Set $\gamma = \frac{1}{\sqrt{1-v^2}}$ and $c = 1$ to get

$$\begin{gathered} t' = \gamma(t - vx) \\ x' = \gamma(x - vt) \end{gathered}$$

1.2.4 4-Vectors

4-vectors are located at a single point in spacetime, not stretched from A to B. At each point p, the set of al possible vectors located at p is T_p, the tangent space of p. The set of all tangent spaces of an n-dimensional manifold M can be assembled into a $2n$-dimensional manifold $T(M)$, the "tangent bundle."



Tangent Space and Manifold Diagram (it doesn't exactly look like this though)

A vector A can be represented as a linear combination of basis vectors $\hat{e}_{(\mu)}$ and coefficients A^μ:

$$A = A^\mu \hat{e}_{(\mu)}$$

The basis vectors transform as expected with Lorentz transformations:

$$\hat{e}_{(\mu)} = \Lambda^{\nu'}_{\mu} \hat{e}_{(\nu')}$$

A tangent vector $V(\lambda)$ to a curve in spacetime parametrized by $x^\mu(\lambda)$ can be represented

$$V^\mu = \frac{dx^\mu}{d\lambda} \qquad \text{(components)}$$
$$V = V^\mu \hat{e}_{(\mu)}$$

1.2.5 Dual Vectors

Kind of a complement to a vector space. The dual vector of the tangent space T_p is denoted T_p^*, the "cotangent space." A vector ω is in T_p^* iff:

$$\omega(aV + bW) = a\omega(V) + b\omega(W) \in \mathbb{R}$$
$$V, W \in T_p$$

Essentially, the dual space is the space of all linear maps from T_p to $\mathbb{R}$. A quick and dirty way to think about it: upper indices denote vectors, lower indices denote dual vectors.

Examples of dual vectors include the gradient, the row vector (corresponding to column vectors), etc.

We can define a "basis dual vector" $\hat{\theta}^{(\nu)}$ in order to represent the dual vector ω:

$$\hat{\theta}^{(\nu)}\left(\hat{e}_{(\mu)}\right) = \delta^\nu_\mu$$
$$\omega = \omega_\mu \hat{\theta}^{(\mu)}$$

Then we can find the effect of a dual vector on a vector:

$$\begin{aligned}\omega(V) &= \omega_\mu \hat{\theta}^{(\mu)}\left(V^\nu \hat{e}_{(\nu)}\right)\\ &= \omega_\mu V^\nu \hat{\theta}^{(\mu)}\left(\hat{e}_{(\nu)}\right)\\ &= \omega_\mu V^\nu \delta^\mu_\nu\\ &= \omega_\mu V^\mu \in \mathbb{R}\end{aligned}$$

A similar calculation shows that vectors are also linear maps on dual vectors, which is a nice symmetry.

$$V(\omega) = \omega(V) = \omega_\mu V^\mu \in \mathbb{R}$$

A simple example of a dual vector in spacetime is the gradient of a scalar function. We denote $d\phi$ as the set of partial derivatives w.r.t. spacetime coordinates.

$$d\phi = \frac{\partial \phi}{\partial x^\mu}\hat{\theta}^{(\mu)}$$

The chain rule is intuitive.

$$\frac{\partial \phi}{\partial x^{\mu'}} = \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial \phi}{\partial x^\mu} = \Lambda^\mu_{\mu'}\frac{\partial \phi}{\partial x^\mu}$$

We define the shorthand $\partial_\mu \phi = \frac{\partial \phi}{\partial x^\mu}$. The gradient of a tangent vector along the parametrized curve $x^\mu(\lambda)$ is

$$\frac{d\phi}{d\lambda} = \partial_\mu \phi \frac{\partial x^\mu}{\partial \lambda}$$

1.3 Tensors

Tensors are a generalization of vectors and dual vectors. Recall that a dual vector is a map from vectors to $\mathbb{R}$. Tensors are a map from a collection of vectors & dual vectors to $\mathbb{R}$. For a tensor of rank (k, l), we map the vectors in T_p^* and dual vectors in T_p to $\mathbb{R}$.

$$T : \underbrace{T_p^* \times \cdots \times T_p^*}_{\text{k terms}} \times \underbrace{T_p \times \cdots \times T_p}_{\text{l terms}} \longrightarrow \mathbb{R}$$

$\times$ is the Cartesian product. $T_p \times T_p$ is the space of the ordered pair of vectors.

A tensor is *multilinear*, meaning it acts linearly in each argument. So for example, a tensor of rank $(1, 1)$ would act like

$$T(a\omega + b\eta, cV + dW) = acT(\omega, V) + adT(\omega, W) + bcT(\eta, V) + bdT(\eta, W)$$

We can further say that a scalar is a tensor of rank (0, 0); a vector is a tensor of rank (1, 0); and a dual vector is a tensor of rank (0, 1). The space of all tensors of rank (k, l) form a vector space, where we can define the *tensor product* $\otimes$ as

$$\begin{gathered}T : \text{tensor of rank } (k, l)\\ S : \text{tensor of rank } (m, n)\\ T \otimes S\left(\omega^{(1)}, \cdots, \omega^{(k)}, \cdots, \omega^{(k+m)}, V^{(1)}, \cdots, V^{(l)}, \cdots, V^{(l+n)}\right)\\ = T\left(\omega^{(1)}, \cdots, \omega^{(k)}, \cdots, \omega^{(k+m)}\right) \times\\ S\left(V^{(1)}, \cdots, V^{(l)}, \cdots, V^{(l+n)}\right)\end{gathered}$$

$T \otimes S$ is a rank $(k+m, l+n)$ tensor.

In spacetime, for a rank (k, l) tensor, there's 4^{k+l} basis tensors. We can represent this simply by summing over each bases.

$$T = T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} \hat{e}_{(\mu_1)} \otimes \cdots \otimes \hat{e}_{(\mu_k)} \otimes \hat{\theta}^{(\nu_1)} \otimes \cdots \otimes \hat{\theta}^{(\nu_k)}$$

Similarly, the action of tensors on a st of vectors and dual vectors is given below. Note that each a (k, l) tensor has k upper and l lower indices. Upper indices transform like vectors, while lower indices transform like dual vectors.

$$T\left(\omega^{(1)}, \cdots, \omega^{(k)}, V^{(1)}, \cdots, V^{(l)}\right) = T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}\ \omega^{(1)}_{\mu_1} \cdots \omega^{(k)}_{\mu_k} V^{(1)\nu_1} \cdots V^{(l)\nu_l}$$

Tensors transform as expected under Lorentz:

$$T^{\mu'_1 \cdots \mu'_k}{}_{\nu'_1 \cdots \nu'_l} = \Lambda^{\mu'_1}{}_{\mu_1} \cdots \Lambda^{\mu'_k}{}_{\mu_k} \Lambda^{\nu_1}{}_{\nu'_1} \cdots \Lambda^{\nu_l}{}_{\nu'_l}\ T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}$$

We're also under no obligation to transform every component. For example, we could just transform a single component of a (1, 1) tensor to make it a vector $\to$ vector map.

$$T^{\mu}{}_{\nu} : V^{\nu} \longrightarrow T^{\mu}{}_{\nu} V^{\nu}$$

It gets complicated, but just try to keep track of the indices. Also note that tensors $\neq$ vectors. Tensors are geometric entities independent of the coordinate system.

1.3.1 Examples of Tensors

Metric The $\eta_{\mu\nu}$ is a (0, 2) tensor. Its action on two vectors is known as the "inner product":

$$\eta(V, W) = \eta_{\mu\nu} V^{\mu} W^{\nu} = V \cdot W$$

If the inner product vanishes, the two vectors are orthogonal. Note that the inner product is a scalar, thus invariant under Lorentz transformations. So a set of orthogonal vectors in one inertial Cartesian frame is also orthogonal after some Lorentz transforms, even though they may not look like it.

The norm of a vector is its inner product with itself:

$$\eta(V, V) = \eta_{\mu\nu} V^{\mu} V^{\nu} \begin{cases} < 0 & V^{\mu} \text{ is timelike} \\ = 0 & V^{\mu} \text{ is lightlike/null} \\ > 0 & V^{\mu} \text{ is spacelike} \end{cases}$$

Kronecker Delta The Kronecker delta δ^μ_ρ is another example of a tensor. It's an identity map from a vector to the same vector. Note that we put the indices in the same "column." That's because order doesn't matter, $\delta^\mu_\rho = \delta^\mu{}_\rho = \delta_\rho{}^\mu$.

The Kronecker delta is defined as the metric times its inverse:

$$\delta^\mu_\rho = \eta^{\mu\nu}\eta_{\nu\rho} = \eta_{\rho\nu}\eta^{\nu\mu}$$

Levi-Civita The Levi-Civita tensor has rank $(0,4)$ in spacetime. We define it as

$$\tilde{\epsilon}_{\mu\nu\rho\sigma} = \begin{cases} +1 & \text{if } \mu\nu\rho\sigma \text{ is an even permutation of } 0123 \\ -1 & \text{if } \mu\nu\rho\sigma \text{ is an odd permutation of } 0123 \\ 0 & \text{otherwise} \end{cases}$$

Electromagnetic Field Strength This is a rank $(0,2)$ tensor composed of the E_i and B_i vectors. Note the Latin indices indicating spacelike 1, 2, 3 components. By Lorentz transforming this tensor, we can get the E and B fields in a different reference frame.

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} = -F_{\nu\mu}$$

So for example, the field tensor transforms under a Lorentz boost as

$$F^{\mu'\nu'} = \Lambda^{\mu'}{}_\mu \Lambda^{\nu'}{}_\nu \, F^{\mu\nu}$$

1.3.2 Manipulating Tensors

Contraction A contraction of a tensor reduces its rank from $(k,l) \longrightarrow (k-1, l-1)$. To do this, we sum over the upper and lower indices.

$$S^{\mu\rho}{}_\sigma = T^{\mu\nu\rho}{}_{\sigma\nu}$$

Order matters. In general $T^{\mu\nu\rho}{}_{\sigma\nu} \neq T^{\mu\rho\nu}{}_{\sigma\nu}$.

Raising & Lowering Indices We can also raise and lower indices in tensors using the metric and its inverse. For example,

$$T^{\alpha\beta\mu}{}_\delta = \eta^{\mu\gamma} \, T^{\alpha\beta}{}_{\gamma\delta}$$
$$T_\mu{}^\beta{}_{\gamma\delta} = \eta_{\mu\alpha} \, T^{\alpha\beta}{}_{\gamma\delta}$$

Since the metric and its inverse are, well, inverse of each other, we also have the following:

$$A^\lambda B_\lambda = \eta^{\lambda\rho} A_\rho \eta_{\lambda\sigma} B^\sigma = \delta^\rho_\sigma A_\rho B^\sigma = A_\sigma B^\sigma$$

Symmetric & Antisymmetric Tensors A tensor is called symmetric if it remains unchanged under exchange of indices. In the following, S is symmetric in μ and ν.

$$S_{\mu\nu\rho} = S_{\nu\mu\rho}$$

We can symmetrize a tensor by summing over its permutations. The symmetric part is denoted by parentheses.

$$T_{(\mu_1\mu_2\cdots\mu_n)\rho}{}^{\sigma} = \frac{1}{n!}\left(T_{\mu_1\mu_2\cdots\mu_n\rho}{}^{\sigma} + \text{ sum over permutations of } \mu_1\cdots\mu_n\right)$$

A tensor is called antisymmetric if it changes sign under exchange of indices. In the following, A is symmetric in μ and ρ.

$$A_{\mu\nu\rho} = -A_{\rho\nu\mu}$$

We can antisymmetrize a tensor by taking the alternating sum over its permutations. The antisymmetric part is denoted by a bracket. "Alternating sum" means that the permutations that result from an odd number of exchanges are given a negative sign.

$$T_{[\mu_1\mu_2\cdots\mu_n]\rho}{}^{\sigma} = \frac{1}{n!}\left(T_{\mu_1\mu_2\cdots\mu_n\rho}{}^{\sigma} + \text{ alternating sum over permutations of } \mu_1\cdots\mu_n\right)$$

When contracting over a symmetric tensor, only the symmetric parts matter. Similar for antisymmetric tensors.

$$X^{(\mu\nu)}Y_{\mu\nu} = X^{(\mu\nu)}Y_{(\mu\nu)}$$
$$X^{[\mu\nu]}Y_{\mu\nu} = X^{[\mu\nu]}Y_{[\mu\nu]}$$

We can also decompose a tensor into a sum for *two indices*. For 3 and above, we generally can't.

$$T_{\mu\nu\rho\sigma} = T_{(\mu\nu)\rho\sigma} + T_{[\mu\nu]\rho\sigma}$$

Trace The trace of a tensor, denoted X is given by

$$X = X^{\lambda}{}_{\lambda}$$

That's the simple case of a (1, 1) tensor. For a (0, 2) tensor $Y_{\mu\nu}$, we raise indices first.

$$Y = Y^{\lambda}{}_{\lambda} = \eta^{\mu\nu}Y_{\mu\nu}$$

1.4 Tensor Form of Maxwell's Equations

The classical differential form of Maxwell's Equations:

$$\begin{cases} \nabla \times \mathbf{B} - \partial_t \mathbf{E} = J & (1) \\ \nabla \cdot \mathbf{E} = \rho & (2) \\ \nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0 & (3) \\ \nabla \cdot \mathbf{B} = 0 & (4) \end{cases}$$

We convert to component notation. Spatial indices (Latin i, j, k) are raised and lowered at will since the metric δ_{ij} is an identity on flat 3D space. We also replace J with the current 4-vector $J^\mu = (\rho, J^x, J^y, J^z)$.

$$\begin{cases} \tilde{\epsilon}^{ijk}\partial_j B_k - \partial_0 E^i = J^i & (5) \\ \partial_i E^i = J^0 & (6) \\ \tilde{\epsilon}^{ijk}\partial_j E_k + \partial_0 B^i = 0 & (7) \\ \partial_i B^i = 0 & (8) \end{cases}$$

We use the electromagnetic field strength tensor from before.

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} = -F_{\nu\mu}$$

Now note that $F^{0i} = \eta^{00}\eta^{ii}F_{0i} = E^i$ (we raise and lower indices at will). Also, $F^{ij} = \tilde{\epsilon}^{ijk}B_k$. Thus the first two of Maxwell's equations can be rewritten as

$$\partial_j F^{ij} - \partial_0 F^{0i} = J^i$$
$$\partial_i F^{0i} = J^0$$

We can merge the two equations into one:

$$\partial_\mu F^{\nu\mu} = J^\nu$$

Similarly, the 3rd and 4th of Maxwell's equations can be merged into one, giving us the two equations that describe E and B:

$$\partial_\mu F^{\nu\mu} = J^\nu$$
$$\partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} = 0$$

1.5 Energy & Momentum

Consider the worldline of a single particle. The path is a parametrized curve $x^\mu(\lambda)$ for some variable λ. The tangent vector is $\frac{dx^\mu}{d\lambda}$. If this tangent vector is timelike/spacelike/null at some value of λ, the path is also timelike/spacelike/null at that point.

Consider paths of massive particles. They'll be timelike, slower than c. We'll attempt to parametrize them in terms of their proper time τ (the time seen on the clock of an observer on that path). Recall that

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$$
$$d\tau^2 = -\eta_{\mu\nu} dx^\mu dx^\nu$$

We integrate over these to get the total path:

$$\Delta s = \int \sqrt{\eta_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda$$
$$\Delta \tau = \int \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda$$

We can use the equation above to parametrize $\tau(\lambda)$ in terms of λ. Afterwards, we can invert to parametrize $\lambda(\tau)$ in terms of τ. Great, now we can parametrize the path in terms of τ: $x^\mu(\tau)$.

1.5.1 4-Velocity

We define the tangent vector of the path (parametrized w.r.t τ) as the 4-velocity.

$$U^\mu = \frac{dx^\mu}{d\tau}$$

Note that because $d\tau^2 = -\eta_{\mu\nu} dx^\mu dx^\nu$, the 4-velocity is automatically normalized.

$$\eta_{\mu\nu} U^\mu U^\nu = -1 \qquad \text{(with } c = 1\text{)}$$

Note that the 4-velocity isn't the velocity through space, which can have different magnitudes. It's the velocity through spacetime, with the same magnitude for all observers.

Here we were working with timelike paths. Spacelike paths are weirder because it would mean the observer is travelling "faster than c." For null paths, τ vanishes, so more care is needed.

In the rest frame of a particle, the 4-velocity has components $U^\mu = (1, 0, 0, 0)$ assuming $c = 1$.

1.5.2 4-Momentum

Let the rest mass of a particle be m. Its 4-momentum is defined as

$$p^\mu = mU^\mu$$

The energy of a particle is given by

$$E = p^0$$

For particles at rest, $p^0 = m$. But we were setting $c = 1$, so technically it's $E = p^0 = mc^2$. We see Einstein's famous equation for a particle at rest.

For a moving frame, Lorentz transform the 4-momentum. For example, consider a particle moving in the +x direction with velocity $v = \frac{dx}{dt}$. Keep in mind that $c = 1$, so $\beta = v/c = v$. Get used to it.

$$p^\mu = \begin{pmatrix} \gamma & \gamma v & 0 & 0 \\ \gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma E \\ \gamma v E \\ 0 \\ 0 \end{pmatrix}$$

For small v, we know that $\gamma = (1 - v^2)^{-1/2} \approx 1 + \frac{1}{2}v^2$. Thus $p^0 \approx m + \frac{1}{2}mv^2$ and $p^1 \approx mv$, which is the Newtonian energy and momentum, respectively.

We also see that $p_\mu p^\mu = -m^2$. Tweaking this expression gives the full relativsitic energy equation:

$$E^2 = m^2 + \mathbf{p}^2, \quad \mathbf{p}^2 = \delta_{ij} p^i p^j$$

Classically, Newton's 2nd Law can be written as $\vec{F} = m\vec{a} = \frac{d\vec{p}}{dt}$. Relativistically, we use the 4-vector F^μ:

$$F^\mu = m\frac{d^2 x^\mu}{d\tau^2} = \frac{dp^\mu}{d\tau}$$

For example, the Lorentz force $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$ becomes

$$F^\mu = qU^\mu F_\lambda{}^\mu$$

1.5.3 Stress-Energy Tensor

The 4-momentum works great for the case of 1 particle. For multi-particle systems, we describe the state as a "fluid" – a continuum characterized by density, pressure, etc. Instead of a 4-momentum vector field, we define an energy-momentum tensor (also stress-energy tensor) $T^{\mu\nu}$. This is a $(2, 0)$ symmetric tensor that tells us the system's energy density, pressure, etc.

We can interpret $T^{\mu\nu}$ as "The flux of 4-momentum p^μ across a surface of constant x^ν." From this we can see that

i) T^{00}: "Flux of p^0 (energy) across surface x^0 (time)," i.e. energy density

ii) $T^{0i} = T^{i0}$: Momentum density

iii) $T^{ij}, i \neq j$: "Shearing" force between neighboring elements of fluid.

iv) T^{ii}: Force exerted (per unit area) by fluid in x^i-direction

---

To start off, we consider dust (or "matter" to cosmologists). It's a collection of particles at rest w.r.t each other in flat spacetime. Define the number-flux 4-momentum N^μ:

$$N^\mu = nU^\mu$$

Here n is the number density of particles in the rest frame. Note that N^0 is the number density of particles in any frame, while N^i is the number density of particles in the x^i-direction. Let m be the mass of 1 particle. In the rest frame, the energy density is

$$\rho = mn$$

Note that this is the combination of $N^\mu = (n, 0, 0, 0)$ and $p^\mu = (m, 0, 0, 0)$ – the number-density and 4-momentum. Thus ρ is the $\mu = \nu = 0$ component of the tensor $p \otimes N$. We therefore the define the energy-momentum tensor for dust as

$$T^{\mu\nu}_{\text{dust}} = p^\mu N^\nu = mnU^\mu U^\nu = \rho U^\mu U^\nu$$

Now we generalize for *perfect fluids*: the "fluids" completley specified by the rest-frame energy density ρ and the isotropic rest-frame pressure p.

Due to isotropy, $T^{\mu\nu}$ is diagonal in the rest frame. There's no net flux of any component of momentum in an orthogonal direction. Further, $T^{11} = T^{22} = T^{33}$. Thus the only independent numbers are $\rho = T^{00}$ and $p = T^{ii}$.

It turns out that the general energy-momentum tensor of a perfect fluid in any frame is

$$T^{\mu\nu} = (\rho + p)U^\mu U^\nu + p\eta^{\mu\nu}$$

To determine the evolution of a system, we relate the pressure to density: $p = p(\rho)$. Dust is the special case of $p = 0$. An isotropic photon gas has $p = \frac{1}{3}\rho$.

Note: we're in flat spacetime. Things change with curvature.

Also note that the stress-energy tensor is conserved: the divergence vanishes.

$$\partial_\mu T^{\mu\nu} = 0$$

The $\nu = 0$ case expresses the conservation of energy. Otherwise, $\nu = k$ epresses the conservation of the k-th component of momentum.

1.6 Classical Field Theory

1.6.1 Background

This will be a crude overview, because most of it went over my head.

To go from special relativity (flat spacetime) to general relativity (curvature), we promote the metric tensor $\eta_{\mu\nu}$ to a dynamical tensor field $g_{\mu\nu}(x)$.

Consider a particle in 1D with coordinate q(t)$ in classical mechanics. Let the Lagrangian be $L(q, \dot{q})$. Then we define the action S as

$$S = \int L(q, \dot{q}) \, dt$$

And the Euler-Lagrange equations

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0$$

In field theory, we replace the single coordinate $q(t)$ with a set of spacetime-dependent fields $\phi^i(x^\mu)$. Each ϕ^i is a function on spacetime, and i is an index for each individual field. The action S becomes a functional of these fields (a function of ∞ variables).

The Lagrangian is an integral over the space of Lagrange density $\mathcal{L}$. Here, $\mathcal{L}$ is the function of fields ϕ^i and spacetime derivatives $\partial_\mu \phi^i$:

$$L = \int \mathcal{L}(\phi^i, \partial_\mu \phi^i) \, d^3x$$

The action is the integral of the Lagrangian:

$$S = \int L \, dt = \int \mathcal{L}(\phi^i, \partial_\mu \phi^i) \, d^4x$$

Here we use natural units: $c = 1$, $\hbar = 1$, $k_B = 1$. Thus

$$[\text{energy}] = [\text{mass}] = [\text{length}^{-1}] = [\text{time}^{-1}]$$

Considering infinitesimal variations, and doing lots of calculations I couldn't follow, we arrive at the Euler-Lagrange equations for field theory in flat spacetime:

$$\frac{\delta S}{\delta \phi^i} = \frac{\partial \mathcal{L}}{\partial \phi^i} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^i)}\right) = 0$$

The simplest example of a field is a real scalar field.

$$\phi(x^\mu) : (\text{spacetime}) \longrightarrow \mathbb{R}$$

Upon quantization, "excitations" of a field are observable as particles. Different fields correspond to different particle types.

i) Scalar fields: spinless particles

ii) Vector fields & other tensors: higher-spin particles

iii) Complex-valued fields (2 degrees of freedom): particle + antiparticle

1.6.2 Motion & Properties

For the field ϕ, we can apply classical mechanics:

$$E_{\text{kinetic}} = \frac{1}{2}\dot{\phi}^2$$
$$E_{\text{gradient}} = \frac{1}{2}(\nabla\phi)^2$$
$$E_{\text{potential}} = V(\phi)$$

Potential energy is Lorentz invariant. Kinetic and gradient energy aren't Lorentz invariant by themselves, but together they are:

$$-\frac{1}{2}\eta^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) = \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\nabla\phi)^2$$

Thus our Lagrangian (Lagrange density) is

$$\mathcal{L} = -\frac{1}{2}\eta^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) - V(\phi)$$

Using $\Box \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu$, the d'Alembertian, our equation of motion is

$$\Box\phi - \frac{dV}{d\phi} = 0$$

In flat spacetime, we have $\ddot{\phi} - \nabla^2\phi + \frac{dV}{d\phi} = 0$.

For a simple harmonic oscillator with $V(\phi) = \frac{1}{2}m^2\phi^2$, our equation of motion is $\Box\phi - m^2\phi = 0$. This is the famous Klein-Gordon equation.

Note that the signs in our equation are dependent on the sign convention. In this book, it's $(-, +, +, +)$.

For electromagnetic fields, the relevant field is the vector potential A_μ. A_0 is the electrostatic potential ϕ, while the spacelike components are traditional vector potentials A (e.g. magnetic field is $B = \nabla \times A$). We can express the field strength tensor $F_{\mu\nu}$ as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

This exhibits *gauge invariance*: for a small change $A_\mu \longrightarrow A_\mu + \partial_\mu\lambda(x)$,

$$F_{\mu\nu} \longrightarrow F_{\mu\nu} + \partial_\mu\partial_\nu\lambda - \partial_\nu\partial_\mu\lambda = F_{\mu\nu}$$

Gauge invariance, and the symmetries that come with it, will become important in quantum field theory. But that's way too advanced.

The energy-momentum tensor for scalar field theory is

$$T^{\mu\nu}_{\text{scalar}} = \eta^{\mu\lambda}\eta^{\lambda\sigma}\partial_\lambda\phi\partial_\sigma\phi - \eta^{\mu\nu}\left[\frac{1}{2}\eta^{\lambda\sigma}\partial_\lambda\phi\partial_\sigma\phi + V(\phi)\right]$$

The energy-momentum tensor for an electromagnetic field is

$$T^{\mu\nu}_{\text{EM}} = F^{\mu\nu}F^{\nu}{}_{\lambda} - \frac{1}{4}\eta^{\mu\nu}F^{\lambda\sigma}F_{\lambda\sigma}$$

2 Manifolds (pp. 48 - 92)

2.1 Gravity as Geometry

The field giving rise to gravity is the metric tensor itself (which describes the curvature of spacetime). It's not an additional field propagating through spacetime, which is interesting.

The evidence for this is the *universality* of gravitational interaction: The "Principle of Equivalence." Every object is influenced by gravity in the same way. Compare that to an electromagnetic field: particles with different charges behave differently.

2.1.1 Weak Equivalence Principle

(Form 1) Inertial mass and gravitational mass are equal.

Inertial mass m_i is given by Newton's 2nd law: $\vec{F} = m_i\vec{a}$. Gravitational mass is given by Newton's law of gravitation, which can be restated with the gravitational-potential scalar field Φ: $\vec{F}_g = -m_g\nabla\Phi$. The WEP thus states that

$$m_i = m_g$$
$$\vec{a} = -\nabla\Phi$$

In other words, there's a preferred class of trajectories through spacetime on which unaccelerated particles travel, which we call *inertial trajectories*. This can be seen above, where the acceleration depends only on the gradient of the gravitational potential. By "unaccelerated," we mean the particle is subject only to gravity. The WEP has been verified to high precision by the Eötvös Experiment.

Compare this to electromagnetism, where particles with different charges have different trajectories even in the same field.

2.1.2 Einstein Equivalence Principle

The WEP in another form is that

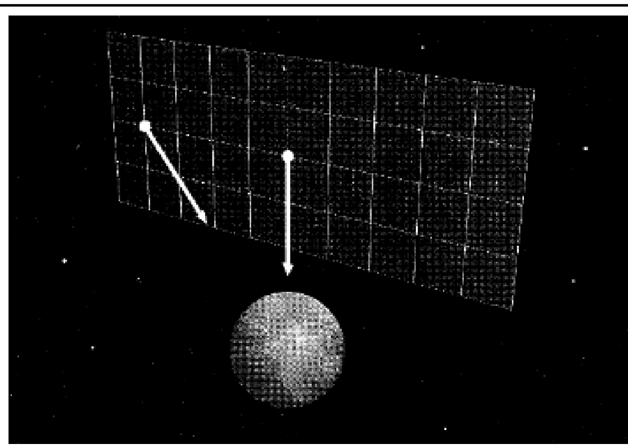
> The motion of freely-falling particles are same in a gravitational field and in a uniformly accelerated frame, in small enough regions of spacetime.

Special relativity showed that mass is just a manifestation of energy. Thus Einstein strengthened the WEP to formulate the Einstein Equivalence Principle (EEP):

> In small enough regions of spacetime, the laws of physics reduce to those of special relativity. It's impossible to detect the existence of gravitational fields by a local experiment.

The EEP accounts for *all* laws of physics, not just free fall. A consequence of the EEP is that "gravitational acceleration" is hard to define: we can't determine the gravitational field by a local experiment. Instead, we let "unaccelerated" mean "freely falling." Thus gravity is not viewed as a force, since no "acceleration" is caused in this view.

We must also give up the notion of inertial frames extending throughout spacetime. We can only have *locally inertial frames.* At some distance away, "acceleration" appears to be observed, i.e. the particle doesn't seem to be stationary in the reference frame:



(Carroll 51) At some distance away, it appears that the left particle is not stationary in the reference frame defined by the right particle.

As an example of the EEP, consider 2 rockets separated by a distance z. Both roeckets accelerate at the same rate a. The trailing rocket emits a photon with wavelength λ_0 at time t_0. We consider the speed of the rockets to be much less than c, $v \ll c$. Then the photon reaches the front rocket after $\Delta t = \frac{z}{c}$. The rockets also pick up $\Delta v = a\Delta t = \frac{az}{c}$.

The light wave should be redshifted by the Doppler effect. Assuming $v \ll c$,

$$\frac{\Delta\lambda}{\lambda_0} \approx \frac{\Delta v}{c} = \frac{az}{c^2}$$

Now let's consider a *gravitational redshift*. Consider a tower at height z on Earth. A photon is emitted from the ground at time t_0. Let $\vec{a_g}$ be the strength of the gravitational field, what we call gravitational acceleration in Newtonian terms. Then by the equivalence principle, there's no difference between this and the rockets. Though the tower seems "stationary" on Earth, it's actually being accelerated against the gravitational field (which pulls it towards the center). The same redshift occurs:

$$\frac{\Delta\lambda}{\lambda_0} \approx \frac{az}{c^2}$$

The redshift is usually stated as $\vec{a_s} = \nabla\Phi$, where Φ is the Newtonian potential and $\vec{a_s}$ is the acceleration of the reference frame (not the particle — hence why the signs are reversed). We combine this with $\frac{\Delta\lambda}{\lambda_0} \approx \frac{\Delta v}{c}$ from before to get

$$\begin{aligned}
\frac{\Delta\lambda}{\lambda_0} &\approx \int \nabla\Phi \; dt \\
&= \frac{1}{c^2}\int \partial_z\Phi \; dz && \text{(reparametrize with z)} \\
&= \Delta\Phi && \text{(with } c = 1\text{)}
\end{aligned}$$

Thus the gravitational redshift depends on the change in gravitational potential. This is because the path in spacetime through which the photons travelled was curved.

2.2 Manifolds

A manifold is a space that may be curved with complicated topology, but locally looks like $\mathbb{R}^n$. For example:

i) $\mathbb{R}^n$ itself (line $\mathbb{R}^1$, plane $\mathbb{R}^2$, etc.)

ii) n-sphere S^n (circle S^1, sphere S^2, etc.)

iii) n-torus T^n (from folding an n-dimensional cube)

iv) Riemann surface of genus g (2-torus with g holes)

Manifolds aren't limited to shapes. For example, $SO(2)$, the set of rotations in 2D, is a manifold. It turns out to be the same manifold as S^1.

The product of 2 manifolds, e.g. $M \times M'$ is also a manifold. If M has dimension n with $p \in M$ and M' has dimension n' with $p' \in M'$, then $M \times M'$ has dimension $n + n'$ with ordered pairs (p, p').

2.2.1 A More Rigorous Definition

Rigor is mathematicians propaganda. But it's helpful to get an idea. First, some basic definitions for maps:

- A map between two sets is a generalized function, denoted $\phi : M \to N$.
- A map is injective (one-to-one) if each element in N has *at most* one lement in M mapped to it.
- A map is surjective (onto) if each element in N has *at least* one lement in M mapped to it.
- If a map is both injective and surjective, it's bijective (invertible) with the inverse map $\phi^{-1} : N \to M$.

For the map $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(x) = e^x$ is injective but not surjective. $\phi(x) = x^3 - x$ is surjective but not injective. $\phi(x) = x^3$ is both.

For the map $\phi : M \to N$,

- M is the domain of the map.
- The image of ϕ is the set of points in N that M gets mapped to.
- The preimage of U under ϕ, denoted $\phi^{-1}(U)$, is the set of elements in M that get mapped to $U \subset N$.

For a general map $\phi : \mathbb{R}^m \to \mathbb{R}^n$, we're mapping the coordinates $(x^1, x^2, \ldots, x^m)$ to $(y^1, y^2, \ldots, y^n)$:

$$\begin{aligned}
y^1 &= \phi^1(x^1, x^2, \ldots, x^m) \\
y^2 &= \phi^2(x^1, x^2, \ldots, x^m) \\
&\vdots \\
y^n &= \phi^n(x^1, x^2, \ldots, x^m)
\end{aligned}$$

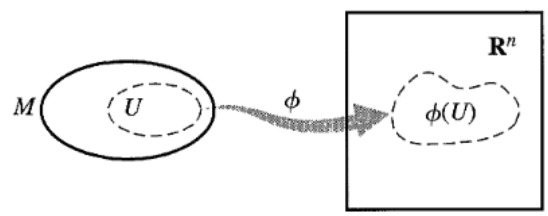
Refer to ϕ as C^p if the p-th derivative exists for all $\phi^1 \ldots \phi^n$. Then a C^0 is map is continuous but not necessarily differentiable, for example. A C^∞ map is both continuous and infinitely differentiable, and we call this "smooth."

Additionally, M, N are *diffeomorphic* if there exists a C^∞ map $\phi : M \to N$ with a C^∞ inverse $\phi^{-1} : N \to M$. Here ϕ is the *diffeomorphism*. This is the mathematical statement of our notion that two spaces are the "same" manifolds. For example, $SO(2)$ is diffeomorphic to S^1.

Now let's start to define open sets:

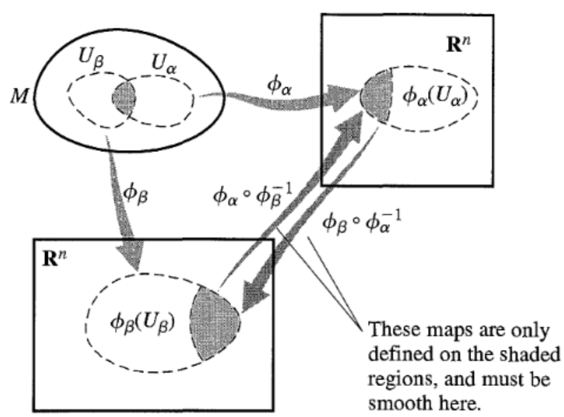
- The *open ball*: a set of points $\{x\} \in \mathbb{R}^n$ such that $|x - y| < r$ for some fixed $y \in \mathbb{R}^n$ and $r \in \mathbb{R}$. Here $|x - y|$ is the distance, $|x - y| = \sqrt{\sum_i (x_i - y_i)^2}$
- The *open set*: a set in $\mathbb{R}^n$ constructed from the arbitrary union of open balls

- I.e. $V \subset \mathbb{R}^n$ is open if $\forall v \in V$, there exists an open ball centered at y that's completely in V.
- Charts/coordinate systems: a subset U of M along with a one-to-one map $\phi : U \to \mathbb{R}^n$ such that the image $\phi(U)$ is open in $\mathbb{R}^n$
  - Note that any map is onto its image, so ϕ is invertible.
  - Then U is an *open set* in M.



Charts/coordinate systems diagram (Carroll p. 59)

- C^∞ atlas: indexed collection of charts $\{(U_\alpha, \phi_\alpha)\}$ such that a) The union of U_α is M, b) The charts are smoothly sewn together.
  - By "smoothly sewn together," we mean that if $U_\alpha \cap U_\beta \neq \emptyset$, the map $(\phi_\alpha \circ \phi_\beta^{-1})$ takes points in $\phi_\beta(U_\alpha \cap U_\beta) \subset \mathbb{R}^n$ *onto* the open set $\phi_\alpha(U_\alpha \cap U_\beta)$, and all of these maps are C^∞ whenever they're defined.



"Smoothly sewn together" diagram (Carroll p. 60)

Now we can finally say: a C^∞ n-dimensional manifold is a set M along with a maximal atlas: one that contains every possible compatible chart.

This is our notion of a set that locally looks like $\mathbb{R}^n$.

Why do we need charts & atlases? Many manifolds can't be covered by a single chart. Consider S^1. The conventional description might be $\theta : S^1 \to \mathbb{R}$ with $\theta \in [0, 2\pi)$. But we require $\theta(S^1)$ to be open in $\mathbb{R}$, so we can't include both $\theta = 0$ and $\theta = 2\pi$. Thus we need at least 2 charts to fully cover S^1.

2.2.2 The Chain Rule

Consider the maps $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^l$. The composition is $(f \circ g) : \mathbb{R}^m \to \mathbb{R}^l$. Lable points x^a on $\mathbb{R}^m$, y^b on $\mathbb{R}^n$, and z^c on $\mathbb{R}^l$.

The chain rule is

$$\frac{\partial}{\partial x^a}(g \circ f)^c = \sum_b \frac{\partial f^b}{\partial x^a}\frac{\partial g^c}{\partial y^b}$$

Its simplified form is

$$\frac{\partial}{\partial x^a} = \sum_b \frac{\partial y^b}{\partial x^a}\frac{\partial}{\partial y^b}$$

2.3 Vectors Revisited

Recall that the tangent space was the set of all vectors at a point in spacetime. Here we revisit that notion a bit more rigorously.

The main problem is, how do we construct a tangent space at a point p in the manifold M *independently of coordinates*?

Let $\mathcal{F}$ be the set of all smooth functions f on M, i.e. the set of all all C^∞ maps $f : M \to \mathbb{R}$. Each curve through p defines the "directional derivative" operator, mapping $f \to \frac{df}{d\lambda}$ at p. Here λ is the curve parametrization.

The tangent space T_p can be identified as the space of directional derivative operators along curves through p.

Proof: First, we have to show that the space of directional derivatives is a vector space. This is straightforward: consider operators $\frac{d}{d\lambda}$ and $\frac{d}{d\eta}$, representing derivatives along $x^\mu(\lambda)$ and $x^\mu(\eta)$ through p. These are obviously linear. They also obey the Leibniz (product) rule:

$$\begin{aligned}\left(a\frac{d}{d\lambda} + b\frac{d}{d\eta}\right)(fg) &= af\frac{dg}{d\lambda} + ag\frac{df}{d\lambda} + bf\frac{dg}{d\eta} + bg\frac{df}{d\eta} \\ &= \left(a\frac{df}{d\lambda} + b\frac{df}{d\eta}\right)g + \left(a\frac{dg}{d\lambda} + b\frac{dg}{d\eta}\right)f\end{aligned}$$

So we're justified in saying that the space of directional derivatives is a vector space.

Now we have to show that this is the vector space we want: it should have the same dimension as M, and yield the natural idea of a vector pointing in a certain direction. The easiest way to do this is by finding a basis for the space. Consider the coordinate chart to be x^μ. This has a set of n directional derivatives at p, the *partial derivatives*.

Partial derivatives: directional derivatives parametrized by x^μ itself along the curve defined by $x^\nu = \text{const.}$, for all $x^\nu \neq x^\mu$.

Our claim is that the partial derivative operators $\{\partial_\mu\}$ at p form a basis for the tangent space T_p.

Proof: It suffice to show that any directional derivative can be decomposed into partial derivatives. Consider the n-manifold M, coordinate chart $\phi : M \to \mathbb{R}^n$, the curve $\gamma : \mathbb{R} \to M$, and a function $f : M \to \mathbb{R}$.

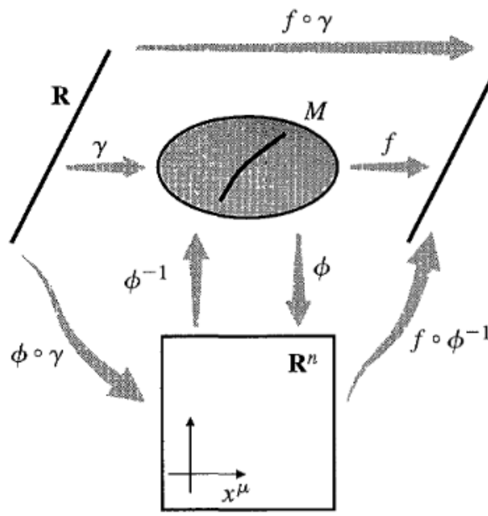


Diagram for above conditions (Carroll 65)

Let λ be the parameter for the curve γ. Then express $\frac{d}{d\lambda}$ in terms of partials:

$$\begin{aligned}
\frac{d}{d\lambda}f &= \frac{d}{d\lambda}(f \circ \gamma) \\
&= \frac{d}{d\lambda}\left[(f \circ \phi^{-1}) \circ (\phi \circ \gamma)\right] \\
&= \frac{d(\phi \circ \gamma)^\mu}{d\lambda}\frac{\partial(f \circ \phi^{-1})^\mu}{\partial x^\mu} && \text{(chain rule)} \\
&= \frac{dx^\mu}{d\lambda}\partial_\mu f
\end{aligned}$$

We've shown that the partials are a good basis for the vector space of directional derivatives. In fact, the particular basis $\hat{e}_{(\mu)} \equiv \partial_\mu$ is the *coordinate basis* for T_p. However, note that for curved manifolds, the coordinate basis vectors usually aren't orthonormal.

The transformation law under a change of coordinates is intuitive. Assume basis vectors $\hat{e}_{(\mu)} \equiv \partial_\mu$. Now the basis vectors become

$$\partial_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu$$

Thus for a general $V = V^\mu \partial_\mu$,

$$V^{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} V^\mu$$

2.3.1 Commutators

A vector is a directional derivative operator at a curve through a point, Thus a *vector field* is a map from smooth functions to smooth functions all over the manifold.

For vector fields X and Y, define the commutator by its action on a function $f(x^\mu)$:

$$[X, Y](f) \equiv X\left(Y(f)\right) - Y\left(X(f)\right)$$

THis is itself a vector field: it's linear and obeys the Leibniz rule.

$$[X, Y](af + bg) = a[X, Y](f) + b[X, Y](g) \tag{linear}$$
$$[X, Y](fg) = f[X, Y](g) + g[X, Y](f) \tag{Leibniz}$$

The specific coordinates are given by

$$[X, Y]^\mu = X^\lambda \partial_\lambda Y^\mu - Y^\lambda \partial_\lambda X^\mu$$

The commutator is sometimes called a *Lie bracket* and it's a special case of the Lie derivative. As partials commute, the commutator of the vector field given by the partial derivatives $\{\partial_\mu\}$ of a coordinate function always vanishes.

2.4 Tensors Revisited

Consider dual vectors (one-forms). The space of dual vectors is the cotangent space T_p^*, or the set of linear maps $\omega : T_p \to \mathbb{R}$. The classic example is the gradient df. Its action on $\frac{d}{d\lambda}$ is

$$df\left(\frac{d}{d\lambda}\right) = \frac{df}{d\lambda}$$

The gradient encodes the necessary information to find the directional derivatives along any curve through $p \in M$, fulfilling the role of a dual vector.

For the tangent space T_p, the partials along the coordinate axes were a natural basis. For the cotangent space T_p^*, the gradients of the coordinate function x^μ are a natural basis. In flat spacetime, we made a basis for T_p^* by requiring that $\hat{\theta}^{(\mu)}(\hat{e}_{(\nu)}) = \delta^\mu_\nu$. Now, we have the condition

$$dx^\mu(\partial_\nu) = \frac{\partial x^\mu}{\partial x^\nu} = \delta^\mu_\nu$$

And an arbitrary dual vector (one-form) can be expanded into its components as $\omega = \omega_\mu dx^\mu$. The transformation laws for tensors are intuitive. For a (k, l) tensor,

$$T^{\mu'_1 \dots \mu'_k}{}_{\nu'_1 \dots \nu'_l} = \frac{\partial x^{\mu'_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x^{\mu'_k}}{\partial x^{\mu_k}} \frac{\partial x^{\nu_1}}{\partial x^{\nu'_1}} \cdots \frac{\partial x^{\nu_l}}{\partial x^{\nu'_l}} T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}$$

Note: the partial derivatives of a higher-rank tensor aren't tensorial, generally. Later, we'll define new tensorial operators — exterior derivatives, covariant derivatives, Lie derivatives.

2.5 The Metric & Canonical Form

The metric is obviously pretty important in relativity. Some uses are (Sachs & Wu 1977):

1. Supplies notion of "past" & "future"
2. Allows computation of path length & proper time
3. Determines "shortest distance" between 2 points
4. Replaces Newtonian gravitational field ϕ
5. Provides notion of locally inertial frames
6. Determines causality by defining speed of light
7. Replaces traditional Euclidiean 3D dot product

Recall that line elements in special relativity are $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$ with the Minkowski metric $\eta_{\mu\nu}$. In general relativity, we use the metric tensor $g_{\mu\nu}$ to write

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

Note 1: dx isn't the "infinitesimal displacement" from calculus, but a rigorous notion of a *basis one-form* given by the gradient function. Generally, we can treat it like an infinitesimal and operate as such. Still good know that it's something else.

Note 2: ds^2 isn't a square of any quantity, nor a displacement. It's just a conventional shorthand for the metric tensor, a map of 2 vectors to one real:

$$g_{\mu\nu} V^\mu W^\nu = g(V, W) = ds^2(V, W)$$

Meanwhile, $(dx)^2$ refers to $dx \otimes dx$, a $(0, 2)$ tensor.

The **canonical form** of $g_{\mu\nu}$ is a diagonal matrix:

$$g_{\mu\nu} = \mathrm{diag}(-1, \dots, -1, +1, \dots, +1, 0, \dots, 0)$$

The **signature** is the number of both positive and negative eigenvalues. For example, Minkowski spacetime has a "-+++" signature.

- If all eigenvalues are > 0, the metric is Euclidean/Riemannian.
- If there's a single -1, the metric is is Lorentzian/pseudo-Riemannian.
- If there's a mix of $+1$ and -1, the metric is indefinite.
- If any eigenvalue is 0, the metric is degenerate.

At any $p \in M$, there exists a coordinate system $x^{\hat{\mu}}$ in which $g_{\hat{\mu}\hat{\nu}}$ has a canonical form, and its first derivatives $\partial_{\hat{\sigma}} g_{\hat{\mu}\hat{\nu}}$ vanish. This is useful because it allows us to treat small enough regions of spacetime as flat.

$$g_{\hat{\mu}\hat{\nu}}(\rho) = \eta_{\hat{\mu}\hat{\nu}} \quad , \quad \partial_{\hat{\sigma}} g_{\hat{\mu}\hat{\nu}}(\rho) = 0$$

We call this coordinate system $x^{\hat{\mu}}$ **locally inert coordinates**. As an example, consider an observer with 4-velocity $U^{\hat{\mu}}$ and a rocket flying past with 4-velocity $V^{\hat{\mu}}$. In special relativity, we can work in the observer's reference frame, in which

$$U^{\hat{\mu}} = (1, 0, 0, 0) \quad , \quad V^{\hat{\mu}} = (\gamma, \gamma\beta, 0, 0)$$

We see that $\gamma = -\eta_{\hat{\mu}\hat{\nu}} U^{\hat{\mu}} V^{\hat{\nu}} = -U_{\hat{\mu}} V^{\hat{\mu}}$ using the -+++ signature for the Minkowski metric. Then, we know that $\gamma = 1/\sqrt{1-\beta^2}$, so $\beta = \sqrt{1-\gamma^{-2}}$. Thus we have

$$\beta = \sqrt{1 - (U_{\hat{\mu}} V^{\hat{\mu}})^{-2}}$$

In curved spacetime, the metric isn't flat. But at the point of measurement, we can use locally inertial coordinates where $g_{\hat{\mu}\hat{\nu}} = \eta_{\hat{\mu}\hat{\nu}}$. Thus the same equation holds in general relativity, which shouldn't be a surprise since the equation is purely tensorial (doesn't depend on coordinates).

2.6 Expanding Universe

Consider the metric below (special case of the Robertson–Walker metric):

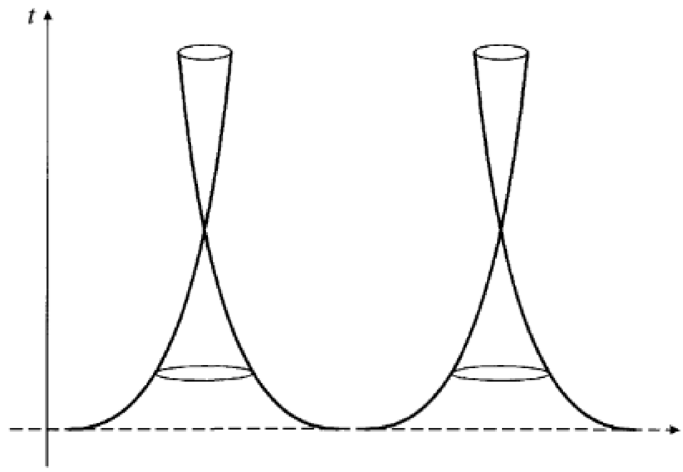
$$ds^2 = -dt^2 + a(t)^2 \left[dx^2 + dy^2 + dz^2\right]$$

This models an expanding universe with scale factor $a(t)$. The solutions for the scale factor can be of the form $a(t) = t^q$, with $0 < q < 1$ and $0 < t < \infty$.

Now consider null paths, where $ds^2 = 0$. Hold y, z as constant. Then

$$\begin{gathered} 0 = -dt^2 + t^{2q} dx^2 \\ \frac{dx}{dt} = \pm t^q \\ t = (1-q)^{1/(1-q)} \left(\pm x - x_0\right)^{1/(1-q)} \end{gathered}$$

Note again, to do this rigorously we should've treated dx as a one-form. But handwaving it with differentials works, and we're physicists not mathematicians.



Spacetime diagram of expanding universe (Carroll 78). Note the singularity at $t = 0$ and two light cones, for two events. Unlike Minkowski spacetime, light cones here need not intersect in the past.

2.7 Causality

2.7.1 Definitions Bash

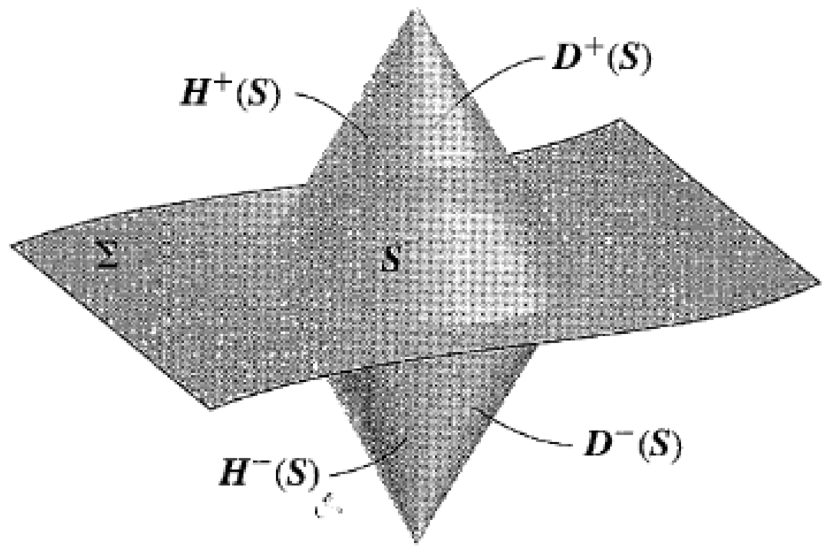
A common question in physics is "given the state at time t_1, can we predict the state at time t_2?" In general relativity, causality must be handled more carefully.

Define a **causal curve** to be one that's timelike or null everywhere. Then for some subset S of the manifold M, define the **causal future** of S, $J^+(S)$, as the set of points that can be reached from S by following a future-directed causal curve. Similarly, $J^-(S)$ is the set of points that cn be reached form S by following a past-directed causal curve.

Define the **chronological future** of S, $I^+(S)$ as the set of points that can be reached from S by following a future-directed *timelike curve* only.

We call a subset $S \subset M$ **achronal** if no two points in S are connected by a timelike curve. Given a closed achronal set S, define the **future domain of dependence** of S, $D^+(S)$, as the set of all points p such that every past-moving inextendible causal curve through p intersects S. The boundary of $D^+(S)$ is the **future Cauchy horizon**, $H^+(S)$. This is a null surface. We define the past domain of dependence $D^-(S)$ and the past Cauchy horizon $D^-(S)$ similarly.

That's difficult to understand, so here's a diagram:



Future domain of dependence and Cauchy horizon on S

The **domain of dependence** is the set of all points for which we can predict what happens by knowing what happens on S:

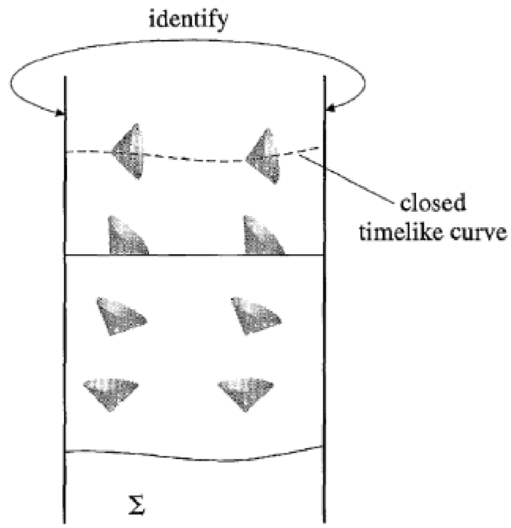
$$D(S) = D^+(S) \cup D^-(S)$$

A closed achronal surface Σ is a **Cauchy surface** if $D(\Sigma)$ is the entire manifold. From information on the Cauchy surface alone, we can predict anything on the manifold. If some region of spacetime has a Cauchy surface, it's **globally hyperbolic** on that region.

A set Σ that's closed, achronal, and has no edge is a **partial Cauchy surface**. This could fail to be a Cauchy surface by its own fault or by some quirk of the spacetime region. For example, singularities will affect how the domain of dependence looks.

2.7.2 Closed Timelike Curves

Now this is an interesting result. In special relativity, an observer is restricted to their own light cone. In general relativity, that becomes a local notion. The curvature of spacetime might "tilt" light cones. If sufficiently distorted, it's possible for an observer to move on a timelike path but intersect himself at a point in the "past."



Closed timlike curves (see the top)

2.8 Tensor Densities

Recall the Levi-Civita symbol:

$$\tilde{\epsilon}_{\mu_1\mu_2\ldots\mu_n} = \begin{cases} +1 & \text{If } \mu_1\mu_2\ldots\mu_n \text{ is an even permutation of } 0,1,\ldots,n-1 \\ -1 & \text{If } \mu_1\mu_2\ldots\mu_n \text{ is an odd permutation of } 0,1,\ldots,n-1 \\ 0 & \text{otherwise} \end{cases}$$

The Levi-Civita symbol isn't a tensor because it's defined not to change under coordinate transformations. But its behavior can be made tensor-like. Note that given an $n \times n$ matrix $M^{\mu}{}_{\mu'}$, the determinant obeys

$$\tilde{\epsilon}_{\mu_1'\mu_2'\ldots\mu_n'}|M| = \tilde{\epsilon}_{\mu_1\mu_2\ldots\mu_n} M^{\mu_1}{}_{\mu_1'} M^{\mu_2}{}_{\mu_2'} \ldots M^{\mu_n}{}_{\mu_n'}$$

By setting $M^{\mu}{}_{\mu'} = \partial x^{\mu}/\partial x^{\mu'}$, we get

$$\tilde{\epsilon}_{\mu_1'\mu_2'\ldots\mu_n'} = \left|\frac{\partial x^{\mu'}}{\partial x^{\mu}}\right| \tilde{\epsilon}_{\mu_1\mu_2\ldots\mu_n} \frac{\partial x^{\mu_1}}{\partial x^{\mu_1'}} \frac{\partial x^{\mu_2}}{\partial x^{\mu_2'}} \ldots \frac{\partial x^{\mu_n}}{\partial x^{\mu_n'}}$$

The Levi-Civita symbol transforms like a tensor, except for the determinant in front. Objects transforming like these are called *tensor densities*. The power to which the Jacobian is raised is called the *weight* of the tensor density.

To convert tensor densities to tensors, multiply by $|g|^{w/2}$, where w is the weight of the density. Thus, the *Levi-Civita tensor* can be defined as

$$\epsilon_{\mu_1\mu_2\ldots\mu_n} = \sqrt{|g|}\tilde{\epsilon}_{\mu_1\mu_2\ldots\mu_n}$$

We can raise indices on this and so on.

2.9 Differential Forms

These are a special kind of tensors. The **differential p-form** is a $(0, p)$ tensor, completely antisymmetric. For example, scalars are 0-forms and dual vectors are 1-forms (hence why we called them “one-forms” earlier).

The space of all p-forms is Λ^p. The space of all p-form fields over a manifold M is $\Lambda^p(M)$.

The basic idea is that differential forms can be differentiated/integrated without the help of additional geometric structures. That’ll be clarified later, hopefully.

2.9.1 Wedge Product

Given a p-form A and q-form B, we can form the $(p+q)$-form $A \wedge B$, the **wedge product** of A and B.

$$(A \wedge B)_{\mu_1 \ldots \mu_{p+q}} = \frac{(p+q)!}{p!q!} A_{[\mu_1 \ldots \mu_p} B_{\mu_{p+1} \ldots \mu_{p+q}]}$$

For example, for 1-forms,

$$(A \wedge B)_{\mu\nu} = 2A_{[\mu} B_{\nu]} = A_\mu B_\nu - A_\nu B_\mu$$

2.9.2 Exterior Derivative

This allows us to differentiate p-forms to get $(p+1)$-form fields. Define the exterior derivative as

$$(dA)_{\mu_1 \ldots \mu_{p+1}} = (p+1) \partial_{[\mu_1} A_{\mu_2 \ldots \mu_{p+1}]}$$

For example, the gradient is the exterior derivative of a 1-form:

$$(d\phi)_\mu = \partial_\mu \phi$$

The exterior derivative satisfies a modified Leibniz rule. For p-form ω and q-form η, we have

$$d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^p \omega \wedge (d\eta)$$

Above, $d(\omega \wedge \eta)$ is a tensor even in curved spacetime. This is in contrast to partial derivatives, which generally aren’t. But exterior derivatives don’t cover teh full range of partial derivatives. They’re only defined for differential p-forms.

We call a p-form **closed** if $dA = 0$ and **exact** if $A = dB$ for some $(p-1)$-form B.

Exterior derivatives have the interesting property that for any differential form A, $d(dA) = 0$. This is sometimes written as $d^2 = 0$.

2.9.3 Hodge Duality

The Hodge star operator is an n-dimensional manifold defined on differential forms. It maps p-forms to $(n-p)$-forms:

$$(*A)_{\mu_1\ldots\mu_{n-p}} = \frac{1}{p!}\epsilon^{\nu_1\ldots\nu_p}{}_{\mu_1\ldots\mu_{n-p}}A_{\nu_2\ldots\nu_p}$$
$$*(*A) = (-1)^{s+p(n-p)}A \ , \quad \text{s} = \# \text{ of negative eigenvalues in metric}$$

In 3D Euclidean space, the Hodge dual of the weidge product of 2 1-forms gives another 1-form. This is the conventional cross product, and only works in 3D.

$$*(U \wedge V)_i = \epsilon_i{}^{jk}U_jV_k$$

Recall the electrodynamic tensor $F : F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}$ \$

And the tensor form of Maxwell's equations:

$$\begin{cases} \partial_\mu F^{\nu\mu} = J^\nu \\ \partial_{[\mu}F_{\nu\lambda]} = 0 \end{cases}$$

From the definition of the exterior derivative, we see that $dF = 0$. Thus there's a 1-form A_μ such that $F = dA$. Here, A is the electromagnetic vector potential.

The other of Maxwell's equations is $d(*F) = *J$. The one-form J is the current 4-vector with lowered indices. If we set $J_\mu = 0$, the equations are invariant under "duality transformations":

$$F \to *F$$
$$*F \to -F$$

Thus the vacuum Maxwell's equations are duality-invariant. But in the presence of charges, the invariancy is spoiled.

2.10 Integration

Tensor densities and differential forms allow us to integate on manifolds. Recall that in ordinary calculus (in $\mathbb{R}$), the "volume element" d^nx picks up a Jacobian under coordinate transformations:

$$d^nx' = \left|\frac{\partial x^{\mu'}}{\partial x^\mu}\right| d^nx$$

Mathematically, this is because on an n-dimensional manifold, the integrand is understood as an n-form. That doesn't make a lot of sense yet. But for an n-dimensional region $\Sigma \subset M$ on the manifold M, the integral can be regarded as a map of n-form fields to real numbers.

$$\int_\Sigma : \omega \to \mathbb{R}$$

For example, in 1 dimension, the one-form is $\omega = \omega(x)dx$. The line integral of this does return a number in $\mathbb{R}$. Here, dx is a differential form, though it helps to handwave it as "infinitesimal distance" in calculations.

Why is the integrand an n-form? To be an n-form, it needs to be an antisymmetric $(0, n)$ tensor. First, write the integral as

$$\int f(x)d\mu \qquad \begin{cases} f(x) : \text{scalar function on manifold} \\ d\mu : \text{volume element ("measure")} \end{cases}$$

The volume element assigns a real number to every infinitesimal region, given n vectors: $d\mu(U_1, U_2, \ldots, U_n) \in \mathbb{R}$. Addition and scalar multiplication work out as expected for vectors and thus tensors. Furthermore, $d\mu$ is antisymmetric. If two vectors are switched, the sign flips.

So we can consider the integrand to be an n-form.

To be more precise, the volume element $d\mu \equiv d^n x$ is an antisymmetric **tensor density** constructed from wedge products.

$$d^n x = dx^0 \wedge \cdots \wedge dx^{n-1}$$

The right-hand side is a coordinate-dependent quanitity, because it depends on the coordinate functions themselves. This is a pretty subtle result, but it's the reason $d^n x$ is considered a tensor density, not a tensor.

To see why $d^n x$ transforms like a tensor density, note the following:

$$\begin{aligned} &1)\ dx^0 \wedge \cdots \wedge dx^{n-1} = \frac{1}{n!}\tilde{\epsilon}_{\mu_1 \ldots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n} \\ &2)\ \tilde{\epsilon}_{\mu_1 \ldots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n} = \tilde{\epsilon}_{\mu'_1 \ldots \mu'_n} \frac{\partial x^{\mu_1}}{\partial x^{\mu'_1}} \cdots \frac{\partial x^{\mu_n}}{\partial x^{\mu'_n}} dx^{\mu'_1} \wedge \cdots \wedge dx^{\mu'_n} \\ &\qquad = \left| \frac{\partial x^{\mu}}{\partial x^{\mu'}} \right| \tilde{\epsilon}_{\mu'_1 \ldots \mu'_n} dx^{\mu'_1} \wedge \cdots \wedge dx^{\mu'_n} \end{aligned}$$

Combining the two and cleaning up any coefficients, we see that

$$d^n x' = \left| \frac{\partial x^{\mu'}}{\partial x^{\mu}} \right| d^n x$$

Thus the volume element $d^n x$ transforms like a tensor density.

Recall that we can form tensors from tensor densities by multiplying with $|g|^{w/2}$, where w is the weight of the tensor. Here, that's just 1. So we can create an invariant volume element by multiplying with $\sqrt{|g|}$:

$$\begin{aligned}\sqrt{|g'|}dx^{0'} \wedge \cdots \wedge dx^{(n-1)'} &= \sqrt{|g|}dx^0 \wedge \cdots \wedge dx^{(n-1)} \\ &= \frac{1}{n!}\epsilon_{\mu_1 \dots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n}\end{aligned}$$

Thus the integral of a scalar function ϕ over an n-dimensional manifold is

$$\int \phi(x)\sqrt{|g|}d^n x$$