

Lecture 3: Tensors and Covariant Derivatives

There is an “old” way and “new” way of defining vectors. Older books defined a contravariant vector as a set of numbers $V_\mu = \{V^1, V^2 \dots V^N\}$ that transform under a change of basis vectors in a particular way. Similarly V^μ was called a covariant vector. It was quite confusing - were they actually two different vectors? We will explore both ways and soon realize that the old way is okay but the new way gives one a much better “feel”.

Contravariant Vectors

- **Old Way:** In the old way, vectors are introduced simply as arrays of numbers that transform in a specific way under a change of coordinates. If we change our coordinates from x^μ to x'^ν , the chain rule from multivariable calculus dictates exactly how our partial derivatives must transform:

$$\frac{\partial}{\partial x'^\nu} = \frac{\partial x^\mu}{\partial x'^\nu} \frac{\partial}{\partial x^\mu} \implies \partial'_\nu = \frac{\partial x^\mu}{\partial x'^\nu} \partial_\mu. \quad (1)$$

This leads to a general definition: we say that V_μ transforms **covariantly** if,

$$V'_\nu = \frac{\partial x^\mu}{\partial x'^\nu} V_\mu. \quad (\text{i.e exactly as } \partial_\mu \text{ transforms}) \quad (2)$$

This gives us yet another concept: we say that V^μ transforms **contravariantly** when,

$$V'^\nu = \frac{\partial x'^\nu}{\partial x^\mu} V^\mu. \quad (3)$$

- **New Way:** We say that at every point p on a manifold $\mathcal{M}$, there exists a **tangent space** $T_p\mathcal{M}$ spanned by a set of basis vectors $\mathbf{e}_\mu$. Let $\mathbf{V}$ be a physical, geometric vector. Because it is a real geometric object, it must exist independently of any chosen coordinate system:

$$\mathbf{V} = V^\mu \mathbf{e}_\mu. \quad (4)$$

To connect this abstract geometry to calculus, we naturally identify these basis vectors with the directional derivative operators along the coordinate curves. That is, we define $\mathbf{e}_\mu \equiv \partial_\mu = \frac{\partial}{\partial x^\mu}$.

Because the vector $\mathbf{V}$ itself is invariant, expressing it as a linear combination in the new primed coordinates must yield the exact same mathematical object as in the unprimed coordinates ($\mathbf{V} = V'^\nu \mathbf{e}'_\nu = V^\mu \mathbf{e}_\mu$). Using our identification with partial derivatives, the chain rule dictates how the basis vectors transform ($\partial'_\nu = \frac{\partial x^\mu}{\partial x'^\nu} \partial_\mu$), allowing us to explicitly show this invariance:

$$V'^\nu \mathbf{e}'_\nu = V'^\nu \partial'_\nu = \frac{\partial x'^\nu}{\partial x^\mu} V^\mu \partial'_\nu = \frac{\partial x'^\nu}{\partial x^\mu} \frac{\partial x^\rho}{\partial x'^\nu} V^\mu \partial_\rho = \delta_\mu^\rho V^\mu \partial_\rho = V^\mu \partial_\mu = V^\mu \mathbf{e}_\mu = \mathbf{V} \quad (5)$$

Bottom line: Contravariant vectors must be thought of as vectors that naturally “live” in the tangent space. They are expressed as a linear combination of basis vectors, i.e. $\mathbf{V} = V^\mu \mathbf{e}_\mu$.

The term “contravariant” arises because the physical vector itself remains geometrically invariant, i.e its components transform oppositely (or “contrary”) to the basis vectors $\mathbf{e}_\mu$.

The Cotangent Space ($T_p^*\mathcal{M}$) and Covariant Vectors

Physics also requires objects that map a tangent vector to a real number, such as the gradient of a scalar potential field ($\partial_\mu \phi$). These objects live in the dual space to the tangent space, known as the **cotangent space** $T_p^*\mathcal{M}$. Elements of this space are called **covariant vectors** (or 1-forms), and their components are written with lower indices, such as ω_μ .

Just as tangent vectors have a natural coordinate basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, covectors have a natural basis. We denote these basis covectors as $\{dx^1, \dots, dx^n\}$. These basis covectors are defined by a very specific rule: dx^μ is the linear machine that measures exactly the x^μ -component of whatever vector you feed it. Mathematically, it is evaluated using the Kronecker delta:

$$dx^\mu(\mathbf{e}_\nu) = dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta_\nu^\mu = \begin{cases} 1 & \text{if } \mu = \nu \\ 0 & \text{if } \mu \neq \nu \end{cases} \quad (6)$$

Any 1-form ω can be written as a linear combination of this basis: $\omega = \omega_\mu dx^\mu$. (Note that 1-form components have *lower* indices, while vector components V^μ have *upper* indices).

Covariant vectors transform “co-variantly” (in the same mathematical manner) as the coordinate basis vectors of the tangent space. If we shift from an unprimed coordinate system x^μ to a primed coordinate system x'^ν , the components of a covariant vector transform according to the chain rule:

$$\omega'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} \omega_\nu \quad (7)$$

The Inner Product on the Tangent Space

Before constructing the metric tensor, we must formally define how to measure lengths and angles in a curved space. In flat Cartesian geometry, we rely on the standard dot product. On a manifold $\mathcal{M}$, we generalize this concept by defining an inner product strictly within the local tangent space $T_p\mathcal{M}$.

An inner product is an abstract mathematical machine, typically denoted by a dot $\cdot$, that takes two tangent vectors $\mathbf{U}, \mathbf{V} \in T_p\mathcal{M}$ and maps them to a real number:

$$\mathbf{U} \cdot \mathbf{V} \in \mathbb{R} \quad (8)$$

To qualify as a valid inner product, this machine must satisfy three strict geometric rules:

1. **Symmetry:** The order of the vectors does not matter. $\mathbf{U} \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{U}$.
2. **Bilinearity:** The machine is linear in both arguments. Scaling a vector scales the output, and the product distributes smoothly over vector addition. $(a\mathbf{U} + b\mathbf{V}) \cdot \mathbf{W} = a(\mathbf{U} \cdot \mathbf{W}) + b(\mathbf{V} \cdot \mathbf{W})$.
3. **Non-degeneracy:** The only vector that yields zero when dot-producted with all other vectors is the zero vector itself.

(Note: In standard Riemannian geometry, we additionally require the inner product to be positive-definite, meaning $\mathbf{V} \cdot \mathbf{V} \geq 0$ for all vectors, equaling zero only if $\mathbf{V}$ is the zero vector. In General Relativity, we relax this to allow for a Lorentzian signature, accommodating light-like and space-like vectors).

Because the inner product is strictly bilinear, its action on any two arbitrary vectors is completely determined by how it acts on the basis vectors $\mathbf{e}_\mu$. We define the components of the metric tensor $g_{\mu\nu}$ as the inner product of the basis vectors themselves:

$$g_{\mu\nu} \equiv \mathbf{e}_\mu \cdot \mathbf{e}_\nu \quad (9)$$

At this juncture, a common point of confusion frequently arises. Consider the 2-sphere (S^2) of radius R . Its coordinate basis vectors are $\mathbf{e}_\theta = \partial_\theta$ and $\mathbf{e}_\phi = \partial_\phi$. Yet, the metric components are known to be $g_{\theta\theta} = R^2$ and $g_{\phi\phi} = R^2 \sin^2 \theta$. How do we reconcile this apparent contradiction if basis vectors are supposed to be unit vectors?

The contradiction stems from mixing the abstract notation of differential geometry with introductory Euclidean concepts. Curvilinear basis vectors are generally *not* normalized to a length of 1. We can resolve this by looking at the geometry from two perspectives:

1. The Extrinsic View (Embedding in 3D Space)

If we embed the sphere in standard Cartesian space, a point on the surface is given by the position vector:

$$\mathbf{r}(\theta, \phi) = \begin{pmatrix} R \sin \theta \cos \phi \\ R \sin \theta \sin \phi \\ R \cos \theta \end{pmatrix} \quad (10)$$

The basis vectors are literally the spatial derivatives of this position vector. Calculating $\mathbf{e}_\theta = \frac{\partial \mathbf{r}}{\partial \theta}$:

$$\mathbf{e}_\theta = \begin{pmatrix} R \cos \theta \cos \phi \\ R \cos \theta \sin \phi \\ -R \sin \theta \end{pmatrix} \quad (11)$$

Taking the standard 3D dot product of this vector with itself yields exactly the metric component:

$$g_{\theta\theta} = \mathbf{e}_\theta \cdot \mathbf{e}_\theta = (R \cos \theta \cos \phi)^2 + (R \cos \theta \sin \phi)^2 + (-R \sin \theta)^2 = R^2(\cos^2 \theta + \sin^2 \theta) = R^2 \quad (12)$$

The same calculation for $\mathbf{e}_\phi = \frac{\partial \mathbf{r}}{\partial \phi}$ naturally yields $g_{\phi\phi} = R^2 \sin^2 \theta$. The vectors stretch and shrink depending on the coordinate location.

2. The Intrinsic View (Abstract Manifold)

In a purely intrinsic view, there is no external 3D space to rely on. The basis vectors $\mathbf{e}_\mu \equiv \partial_\mu$ are abstract directional operators. Here, $g_{\mu\nu} = \mathbf{e}_\mu \cdot \mathbf{e}_\nu$ is not an equation you calculate to *find* the metric; it is the fundamental *definition* of the metric machine. The line element ds^2 directly dictates what these inner products evaluate to.

Therefore, if we express our geometric vectors in component form ($\mathbf{U} = U^\mu \mathbf{e}_\mu$ and $\mathbf{V} = V^\nu \mathbf{e}_\nu$), the bilinearity allows us to factor out the components to evaluate the machine:

$$\mathbf{U} \cdot \mathbf{V} = (U^\mu \mathbf{e}_\mu) \cdot (V^\nu \mathbf{e}_\nu) = U^\mu V^\nu (\mathbf{e}_\mu \cdot \mathbf{e}_\nu) = g_{\mu\nu} U^\mu V^\nu \quad (13)$$

With this established, the squared length of a single tangent vector is rigorously defined as $|\mathbf{V}|^2 = \mathbf{V} \cdot \mathbf{V} = g_{\mu\nu} V^\mu V^\nu$. The angle α between two intersecting vectors is defined exactly as in flat space:

$$\cos \alpha = \frac{\mathbf{U} \cdot \mathbf{V}}{|\mathbf{U}||\mathbf{V}|} = \frac{g_{\mu\nu} U^\mu V^\nu}{\sqrt{g_{\rho\sigma} U^\rho U^\sigma} \sqrt{g_{\alpha\beta} V^\alpha V^\beta}} \quad (14)$$

Example: Angle between vectors on a 2-Sphere (S^2)

Consider the surface of a sphere with radius R , described by the metric $ds^2 = R^2 d\theta^2 + R^2 \sin^2 \theta d\phi^2$. The non-zero metric components are $g_{\theta\theta} = R^2$ and $g_{\phi\phi} = R^2 \sin^2 \theta$.

Suppose at a point $p = (\theta_0, \phi_0)$, we have two tangent vectors. Vector $\mathbf{U}$ points directly South along a line of longitude, so it has components $U^\mu = (1, 0)$. Vector $\mathbf{V}$ points South-East such that its components are $V^\mu = (1, 1)$. Let us calculate the angle α between them.

First, we compute their inner product using the metric at point p :

$$\mathbf{U} \cdot \mathbf{V} = g_{\theta\theta} U^\theta V^\theta + g_{\phi\phi} U^\phi V^\phi = R^2(1)(1) + R^2 \sin^2 \theta_0(0)(1) = R^2 \quad (15)$$

Next, we calculate their respective lengths:

$$|\mathbf{U}| = \sqrt{g_{\theta\theta}(U^\theta)^2 + g_{\phi\phi}(U^\phi)^2} = \sqrt{R^2(1)^2 + 0} = R \quad (16)$$

$$|\mathbf{V}| = \sqrt{g_{\theta\theta}(V^\theta)^2 + g_{\phi\phi}(V^\phi)^2} = \sqrt{R^2(1)^2 + R^2 \sin^2 \theta_0(1)^2} = R\sqrt{1 + \sin^2 \theta_0} \quad (17)$$

Finally, we find the cosine of the angle between them:

$$\cos \alpha = \frac{R^2}{R \cdot R\sqrt{1 + \sin^2 \theta_0}} = \frac{1}{\sqrt{1 + \sin^2 \theta_0}} \quad (18)$$

Notice that the angle physically depends on the latitude θ_0 ! Near the poles where $\sin \theta_0 \rightarrow 0$, the angle approaches 0° (the vectors align). Near the equator where $\sin \theta_0 \rightarrow 1$, the angle approaches $\arccos(1/\sqrt{2}) = 45^\circ$.

Constructing Rank-Two Tensors

Having established the tangent space $T_p\mathcal{M}$, we can now construct more complex geometric objects called **tensors**. For beginning students, the most intuitive way to understand a tensor is as a machine—a *multilinear map*—that takes a certain number of vectors as inputs and produces a real number as an output.

If we want a machine that takes two tangent vectors (say, U and V) and gives us a scalar, we are talking about a **rank-2 tensor**, often denoted as a type $(0, 2)$ tensor. Formally, a rank-2 tensor T at a point p is a map:

$$T : T_p\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R} \quad (19)$$

Because the map is multilinear, the action of the tensor is entirely determined by what it does to the basis vectors. We define the *components* of the tensor by feeding it the partial derivative basis vectors we just established:

$$T_{\mu\nu} = T(e_\mu, e_\nu) = T(\partial_\mu, \partial_\nu) \quad (20)$$

If we feed two arbitrary vectors $U = U^\mu \partial_\mu$ and $V = V^\nu \partial_\nu$ into this tensor, the linearity allows us to pull the components out:

$$T(U, V) = T(U^\mu \partial_\mu, V^\nu \partial_\nu) = U^\mu V^\nu T(\partial_\mu, \partial_\nu) = T_{\mu\nu} U^\mu V^\nu \quad (21)$$

This multilinear machine concept perfectly sets up the most important tensor in Riemannian geometry, which we use to measure distances and angles.

Concept Check 2: If a rank-2 tensor T takes two vectors and maps them to a real number, how would you intuitively describe the action of a rank-3 tensor S of type $(0, 3)$? Write down its definition as a multilinear map.

To see how higher-rank tensors are built, we must first understand the **tensor product**, denoted by the symbol $\otimes$. You can think of the tensor product as a mathematical “glue” that binds smaller linear machines together to form a larger, multilinear machine.

If you have a machine **A** that eats one geometric object to produce a real number, and a machine **B** that eats another object to produce a number, their tensor product $\mathbf{A} \otimes \mathbf{B}$ is a new, combined machine that eats *two* objects. It feeds the first object into **A**, feeds the second object into **B**, and simply multiplies the two resulting numbers together.

Using this operation, we can combine our foundational building blocks—vectors and 1-forms—to construct rank-two tensors. Given two contravariant vectors $\mathbf{V} = V^\mu \mathbf{e}_\mu$ and $\mathbf{W} = W^\nu \mathbf{e}_\nu$, and two covariant 1-forms $\alpha = \alpha_\mu dx^\mu$ and $\beta = \beta_\nu dx^\nu$, we can construct three primary types:

- **Contravariant Tensor (Type 2,0):** Constructed from two vectors.

$$\mathbf{T} = \mathbf{V} \otimes \mathbf{W} = (V^\mu \mathbf{e}_\mu) \otimes (W^\nu \mathbf{e}_\nu) = V^\mu W^\nu (\mathbf{e}_\mu \otimes \mathbf{e}_\nu) \quad (22)$$

Here, $T^{\mu\nu} = V^\mu W^\nu$ are the components, and $\mathbf{e}_\mu \otimes \mathbf{e}_\nu$ forms the natural basis for this tensor space.

- **Covariant Tensor (Type 0,2):** Constructed from two 1-forms. The most famous example in differential geometry is the metric tensor $\mathbf{g}$.

$$\mathbf{S} = \alpha \otimes \beta = (\alpha_\mu dx^\mu) \otimes (\beta_\nu dx^\nu) = \alpha_\mu \beta_\nu (dx^\mu \otimes dx^\nu) \quad (23)$$

The components are $S_{\mu\nu} = \alpha_\mu \beta_\nu$.

- **Mixed Tensor (Type 1,1):** Constructed from one vector and one 1-form. This is mathematically equivalent to a linear transformation (like a standard matrix) that maps a vector to another vector.

$$\mathbf{M} = \mathbf{V} \otimes \beta = (V^\mu \mathbf{e}_\mu) \otimes (\beta_\nu dx^\nu) = V^\mu \beta_\nu (\mathbf{e}_\mu \otimes dx^\nu) \quad (24)$$

The components are $M^\mu_\nu = V^\mu \beta_\nu$.

To solidify this concept, let us examine the metric tensor $\mathbf{g}$ using this formal abstract language. While we often compute with its components $g_{\mu\nu}$ as a matrix, the metric itself is an abstract geometric machine.

It is a symmetric, bilinear map that acts on the local geometry of the manifold. Specifically, it eats two vectors from the tangent space $T_p\mathcal{M}$ and spits out a real number, effectively defining their inner (dot) product. Because it consumes tangent vectors, the metric tensor is constructed from the basis elements of the cotangent space, making it a covariant tensor of rank (0,2).

Using the tensor product of our natural dual basis 1-forms, the abstract metric tensor is written as:

$$\mathbf{g} = g_{\mu\nu} dx^\mu \otimes dx^\nu \quad (25)$$

When we feed this tensor machine two copies of an infinitesimal displacement vector $d\mathbf{s} = dx^\alpha \mathbf{e}_\alpha$, the 1-forms cleanly extract the components, perfectly reproducing the traditional line element:

$$\mathbf{g}(d\mathbf{s}, d\mathbf{s}) = g_{\mu\nu} dx^\mu dx^\nu = ds^2 \quad (26)$$

1 The Covariant Derivative

In standard flat-space calculus using Cartesian coordinates, taking the derivative of a vector field is straightforward: you simply take the partial derivatives of its components. However, to correctly capture changes in tensors across a curved manifold, standard partial derivatives are strictly insufficient.

To see why, let us write a vector field explicitly as a sum of its components and its basis vectors: $\mathbf{V} = V^\nu \mathbf{e}_\nu$. By keeping the notation $\mathbf{e}_\nu$ for the basis vectors, we avoid the dangerous notational trap of confusing the basis vector with the differential operator $\partial_\nu = \frac{\partial}{\partial x^\nu}$.

If we want to find how this geometric vector changes as we move in the x^μ direction, we apply the partial derivative operator ∂_μ . Because both the components *and* the basis vectors can change from point to point, we must rigorously apply the product rule:

$$\partial_\mu \mathbf{V} = \partial_\mu (V^\nu \mathbf{e}_\nu) = (\partial_\mu V^\nu) \mathbf{e}_\nu + V^\nu (\partial_\mu \mathbf{e}_\nu) \quad (27)$$

The final term, $\partial_\mu \mathbf{e}_\nu$, represents the rate of change of the *basis vector itself* as we move along the manifold. In a global Cartesian grid, the basis vectors $\mathbf{e}_\nu$ point in the exact same direction everywhere, meaning their spatial derivative is strictly the zero vector ($\partial_\mu \mathbf{e}_\nu = \mathbf{0}$).

But on a curved manifold (or even in flat space using curvilinear coordinates like polar coordinates), the basis vectors twist, stretch, and turn as you move. We need a way to quantify how these basis vectors change. Since the derivative of a basis vector is itself a vector in the tangent space, it can be

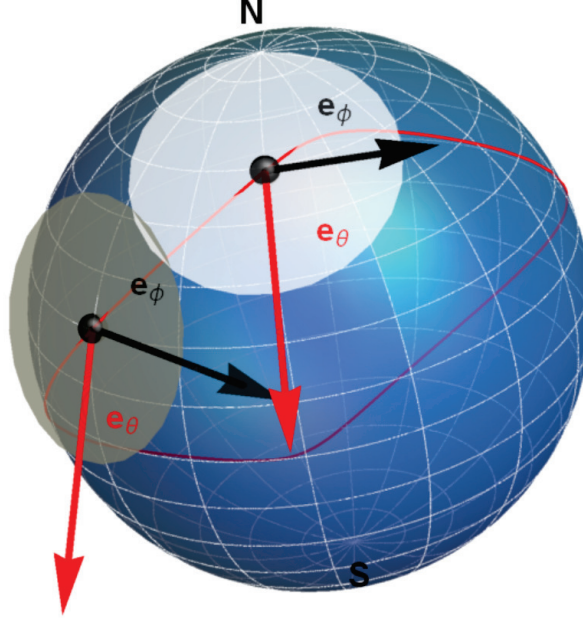


Figure 1: Tangent spaces evaluated at two separated points along a path on the 2-sphere. Because the manifold is curved, the tangent spaces at these distinct locations are distinct planes, and it is crucial to note they are spanned by basis vectors ($\mathbf{e}_\theta$ and $\mathbf{e}_\phi$) that are physically different in orientation and magnitude from one point to the other.

written as a linear combination of the existing basis vectors. The coefficients of this expansion are exactly the **Christoffel symbols** ($\Gamma_{\mu\nu}^\lambda$) that we derived in the previous chapter:

$$\partial_\mu \mathbf{e}_\nu = \Gamma_{\mu\nu}^\lambda \mathbf{e}_\lambda \quad (28)$$

Substituting this geometric relation back into our product rule expression (and swapping the dummy indices ν and λ in the second term so we can factor out the basis vector $\mathbf{e}_\nu$) yields:

$$\partial_\mu \mathbf{V} = (\partial_\mu V^\nu) \mathbf{e}_\nu + V^\lambda (\Gamma_{\mu\lambda}^\nu \mathbf{e}_\nu) = \left(\partial_\mu V^\nu + \Gamma_{\mu\lambda}^\nu V^\lambda \right) \mathbf{e}_\nu \quad (29)$$

The term inside the parentheses represents the true, coordinate-independent rate of change of the vector's components. This necessitates the introduction of a new derivative operator called the **covariant derivative**, denoted by ∇ . For a contravariant vector field V^ν , the covariant derivative is formally defined as the components of this full geometric variation:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma_{\mu\lambda}^\nu V^\lambda \quad (30)$$

The covariant derivative uses the connection coefficients (Christoffel symbols) to properly differentiate vectors, correcting the raw coordinate change (the partial derivative) with a geometric adjustment (how the grid itself bends).

For a covariant vector (a 1-form), the basis transformation applies inversely, so the geometric correction term acquires a minus sign:

$$\nabla_\mu V_\nu = \partial_\mu V_\nu - \Gamma_{\mu\nu}^\lambda V_\lambda \quad (31)$$

This operator guarantees that the derivative of a tensor transforms strictly as a tensor, preserving general coordinate invariance across the manifold.

Physical Example: The Auto-Parallel Pole

Imagine a man walking on a curved manifold along a geodesic curve. He holds a pole of fixed length that lies entirely in the tangent plane and points directly along his path. Is he parallel transporting the pole?

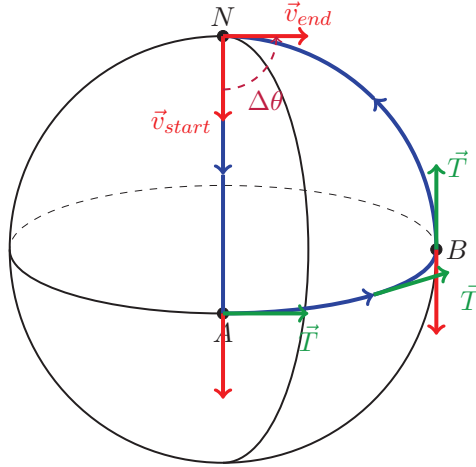
Yes, he is absolutely parallel transporting it. Geometrically, a geodesic is formally defined as an “auto-parallel” curve—meaning it is a curve that parallel transports its own tangent vector. By holding the pole such that it always points along the curve, the man is aligning the pole with the geodesic’s tangent vector. Because the pole has a fixed length, it perfectly corresponds to the tangent vector parameterized by an affine parameter (such as arc length s).

If we denote the geodesic curve as $\gamma(s)$ and the pole as the tangent vector $T = \frac{d\gamma}{ds}$, the condition for the man keeping the pole aligned with the curve without changing its length is strictly equivalent to the geodesic equation:

$$\nabla_T T = 0 \quad (32)$$

Since the covariant derivative of the pole along the path is strictly zero, the pole perfectly satisfies the mathematical requirement for parallel transport.

2 Intrinsic Curvature: Parallel Transport



The most rigorous intrinsic test for curvature is to take a vector, **parallel transport** it along a closed loop, and see if it comes back pointing in the same direction.

Look at the diagram above. Start at the North Pole (N) of a sphere with a vector pointing south along a specific meridian. Parallel transport it down to the equator at point A (keeping the angle between the vector and the path constant). Then, slide it along the equator to point B . Finally, slide it back up to the North Pole.

Because the surface of the sphere is intrinsically curved, the red vector arrives at the North Pole rotated by a measurable angle ($\Delta\theta$) compared to how it started! In a flat space (like a sheet of paper), parallel transporting a vector around any closed loop will *always* result in the exact same vector.

Notice also the green arrows ($\vec{T}$) along the equator; they represent the tangent vector (the auto-parallel pole from our previous example). Unlike the red vector, which points south across the equator, the green pole points directly East, tracking exactly along the geodesic.