Riemannian Geometry: Definitions, Pictures, and Results

Adam Marsh

February 27, 2019

Abstract

A pedagogical but concise overview of Riemannian geometry is provided, in the context of usage in physics. The emphasis is on defining and visualizing concepts and relationships between them, as well as listing common confusions, alternative notations and jargon, and relevant facts and theorems. Special attention is given to detailed figures and geometric viewpoints, some of which would seem to be novel to the literature. Topics are avoided which are well covered in textbooks, such as historical motivations, proofs and derivations, and tools for practical calculations. As much material as possible is developed for manifolds with connection (omitting a metric) to make clear which aspects can be readily generalized to gauge theories. The presentation in most cases does not assume a coordinate frame or zero torsion, and the coordinate-free, tensor, and Cartan formalisms are developed in parallel.

Contents

1 Introduction

Riemannian geometry is fundamental to general relativity, and is also the foundational inspiration for gauge theories. This bifurcation has led to many presentations tending towards either the specific (e.g. presented in tensor notation assuming a coordinate frame and zero torsion) or the abstract (e.g. using the language of fiber bundles). Here we attempt to cover the material in a way that makes clear the relationships between different approaches and notations, while emphasizing intuitive geometric meanings.

In the presentation we try to take an approach which is useful both as a learning tool complementary to other resources, and as a reference which concisely covers the relevant topics. This ends up consisting mainly of clear definitions along with related results. We also attempt to "take pictures seriously," by making explicit the assumptions being made and the quantities being depicted. Thus the three main components are definitions, pictures, and results.

A series of appendices are included which cover relevant material referred to in the presentation. These appendices can either be read before the main presentation or referred to as necessary.

Throughout the paper, warnings concerning a common confusion or easily misunderstood concept are separated from the core material by boxes, as are intuitive interpretations or heuristic views that help in understanding a particular concept. Quantities are written in **bold** when first mentioned or defined.

2 Parallel transport

2.1 The parallel transporter

By definition, for a vector w at a point p of an n-dimensional manifold M, **parallel transport** assigns a vector $\|_C(w)$ at another point q that is dependent upon a specific path C in M from p to q.

To see that this dependence upon the path matches our intuition, we can consider a vector transported in what we might consider to be a "parallel" fashion along the edges of an eighth of a sphere. In this example, the sphere is embedded in $\mathbb{R}^3$ and the concept of "parallel" corresponds to incremental vectors along the path having a projection onto the original tangent plane that is parallel to the original vector.

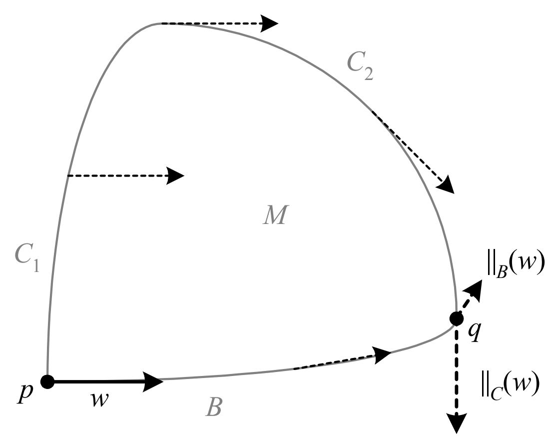

FIGURE 2.1: A vector w transported in what we intuitively consider to be a "parallel" way along two different paths (B and $C = C_1 + C_2$) on a surface results in two different vectors.

The **parallel transporter** is therefore a map

$$\|_C\colon T_pM \to T_qM, \tag{2.1}$$

where C is a curve in M from p to q and T_pM is the tangent space at p (see Section B.2). To match our intuition we also require that this map be linear (i.e. parallel transport is assumed to preserve the vector space structure of the tangent space); that it be the identity for vanishing C; that if $C = C_1 + C_2$ then $\|_C = \|_{C_2}\|_{C_1}$; and that the dependence on C be smooth (this is most easily defined in the context of fiber bundles, which we will not cover here). If we then choose a frame on $U \subset M$, we have bases for each tangent space that provide isomorphisms $T_pU \cong \mathbb{R}^n$, $T_qU \cong \mathbb{R}^n$. Thus the parallel transporter can be viewed as a map

$$\|^{\lambda}{}_{\mu}\colon\ \{C\} \to GL\,(n, \mathbb{R}) \tag{2.2}$$

from the set of curves on U to the Lie group $GL\,(n, \mathbb{R})$ of general linear transformations on $\mathbb{R}^n$; however, it is important to note that the values of $\|^{\lambda}{}_{\mu}$ depend upon the choice of frame.

2.2 The covariant derivative

Having defined the parallel transporter, we can now consider the **covariant derivative**

$$\begin{aligned} \nabla_v w &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(w\,|_{p+\varepsilon v} - \|_C\,(w\,|_p)\right) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\|_{-C}\,(w\,|_{p+\varepsilon v}) - w\,|_p\right), \end{aligned} \tag{2.3}$$

where C is an infinitesimal curve starting at p with tangent v. At a point p, $\nabla_v w$ compares the value of w at $p + \varepsilon v$ to its value at p after being parallel transported to $p + \varepsilon v$, or equivalently in the limit $\varepsilon \to 0$, the value of w at p to its value at $p + \varepsilon v$ after being parallel transported back to p. (see Section B.2 on how $p + \varepsilon v$ is well-defined in the limit $\varepsilon \to 0$).

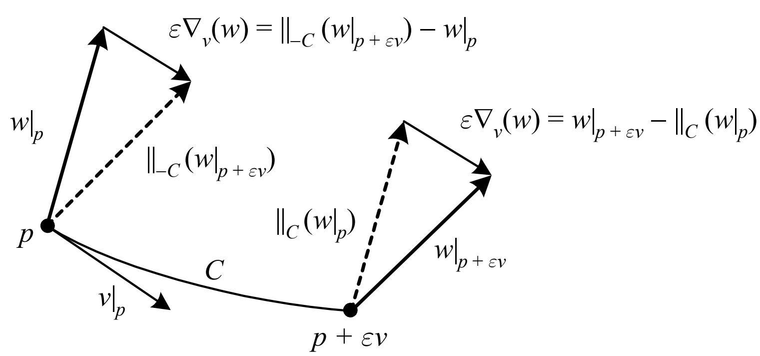

FIGURE 2.2: The covariant derivative $\nabla_v w$ is the difference between a vector field w and its parallel transport in the direction v.

☼ In this and future depictions of vector derivatives, the situation is simplified by focusing on the change in the vector field w while showing the "transport" of w as a parallel displacement. This has the advantage of highlighting the equivalency of defining the derivative at either 0 or ε in the limit $\varepsilon \to 0$. Depicting $\|_C (w|_p)$ as a non-parallel vector at $p+\varepsilon v$ would be more accurate, but would obscure this fact. We also will follow the picture here in using words to characterize derivatives: namely, "the difference" is short for "the difference per unit ε to order ε in the limit $\varepsilon \to 0$."

Two properties of $\nabla_v w$ that are easy to verify are that is is linear in v, and that for a function f on M it obeys the rule

$$\begin{aligned}\nabla_v (fw) &= v(f)\,w + f\nabla_v(w) \\ &= \mathrm{d}f(v)\,w + f\nabla_v(w)\,.\end{aligned} \tag{2.4}$$

As we will see in Section 3.1, this is the Leibniz rule (see Appendix C.1) for the covariant derivative generalized to the tensor algebra. See Section B.6 for a review of the differential d and the relation $v(f) = \mathrm{d}f(v)$. Note that $\nabla_v w$ is a directional derivative, i.e. it depends only upon the value of v at p; v is in effect used only to choose a direction. In contrast, the Lie derivative $L_v w$ (see Section C.2) requires v to be a vector field, since w is in this case compared to its value after being "transported" by the local flow of v, and so depends on the derivative of v at p.

△ It is important to remember that there is no way to "transport" a vector on a manifold without introducing some extra structure.

Instead of parallel transport, one can consider the covariant derivative as the fundamental structure being added to the manifold. In this case it is useful to define the covariant derivative along a smooth parametrized curve $C(t)$ by using the tangent to the curve as the direction, i.e.

$$\frac{\mathrm{D}}{\mathrm{d}t}w \equiv \mathrm{D}_t w \equiv \nabla_{\dot{C}(t)} w, \tag{2.5}$$

where $\dot{C}(t)$ is the tangent to C at t. $\mathrm{D}_t w$ is sometimes called the **absolute derivative** (AKA intrinsic derivative) and its definition only requires that w be defined along the curve $C(t)$. We can then define the parallel transport of $w|_p$ along $C(t)$ as the vector field w that satisfies $\mathrm{D}_t w = 0$.

△ The notation for the absolute derivative is potentially confusing since the implicitly referenced curve $C(t)$ does not appear in the expression $\mathrm{D}_t w$.

2.3 The connection

If we view ∇ as a map from two vector fields v and w to a third vector field $\nabla_v w$, it is called an **affine connection**. Note that since no use has been made of coordinates or frames in the definition of ∇, it is a frame-independent quantity (see Appendix B for a review of coordinates and frames).

Since ∇_v is linear in v, and depends only on its local value, we can regard ∇ as a 1-form on M. If we choose a frame e_μ on M with corresponding dual frame β^μ, we can define the **connection 1-form**

$$\Gamma^\lambda{}_\mu(v) \equiv \beta^\lambda(\nabla_v e_\mu)\,. \tag{2.6}$$

${\Gamma^\lambda}_\mu(v)$ is the λ^{th} component of the difference between the frame e_μ and its parallel transport in the direction v.

From its definition, it is clear that ${\Gamma^\lambda}_\mu$ is a frame-dependent object, and additionally it is not local since it is formed from the derivative of the frame; therefore it cannot be viewed as the components of a tensor (see Appendix A for a review of tensors and forms).

At a point p, the value of ${\Gamma^\lambda}_\mu(v)$ is an infinitesimal linear transformation on T_pM, i.e. ${\Gamma^\lambda}_\mu$ is a frame-dependent 1-form whose values sit in the Lie algebra $gl(n,\mathbb{R})$. Using the notation for algebra- and vector-valued forms defined in Section A.9, we can then write

$$\check{\Gamma}(v)\,\vec{w} \equiv {\Gamma^\lambda}_\mu(v)\,w^\mu e_\lambda = (\nabla_v e_\mu)\,w^\mu, \tag{2.7}$$

where we view $\vec{w}$ as a $\mathbb{R}^n$-valued 0-form. The vector $\check{\Gamma}(v)\,\vec{w}$ measures the difference between the frame and its parallel transport in the direction v, weighted by the components of w.

△ It is important to remember that $\check{\Gamma}(v)\,\vec{w}$ is related to the difference between the frame and its parallel transport, while $\nabla_v w$ measures the difference between w and its parallel transport; thus unlike $\nabla_v w$, $\check{\Gamma}(v)\,\vec{w}$ depends only upon the local value of w, but takes values that are frame-dependent.

△ Since we have used the frame to view $\check{\Gamma}$ as a $gl(n,\mathbb{R})$-valued 1-form, i.e. a matrix-valued 1-form, $\vec{w}$ must be viewed as a frame-dependent column vector of components. We could instead view $\check{\Gamma}$ as a $gl(\mathbb{R}^n)$-valued 1-form and $\vec{w}$ as a frame-independent intrinsic vector. In this case the action of $\check{\Gamma}$ on $\vec{w}$ would be frame-independent, but the value of $\check{\Gamma}$ itself would remain frame-dependent. We choose to use matrix-valued forms due to the need below to take the exterior derivative of component functions, but the abstract viewpoint is important to keep in mind when generalizing to fiber bundles.

2.4 The covariant derivative in terms of the connection

$\nabla_v w$ can be written in terms of $\check{\Gamma}$ by using the Leibniz rule from Section 2.2 with w^μ as frame-dependent functions:

$$\begin{aligned}
\nabla_v w &= \nabla_v (w^\mu e_\mu) \\
&= v(w^\mu)\,e_\mu + w^\mu \nabla_v (e_\mu) \\
&= \mathrm{d}w^\mu(v)\,e_\mu + \check{\Gamma}(v)\,\vec{w} \\
&\equiv \mathrm{d}\vec{w}(v) + \check{\Gamma}(v)\,\vec{w}
\end{aligned} \tag{2.8}$$

Here we again view $\vec{w}$ as a $\mathbb{R}^n$-valued 0-form, so that $\mathrm{d}\vec{w}(v) \equiv \mathrm{d}w^\mu(v)\,e_\mu$. Thus $\mathrm{d}\vec{w}(v)$ is the change in the components of w in the direction v, making it frame-dependent even though w is not. Note that although $\nabla_v w$ is a frame-independent quantity, both terms on the right hand side are frame-dependent. This is depicted in the following figure.

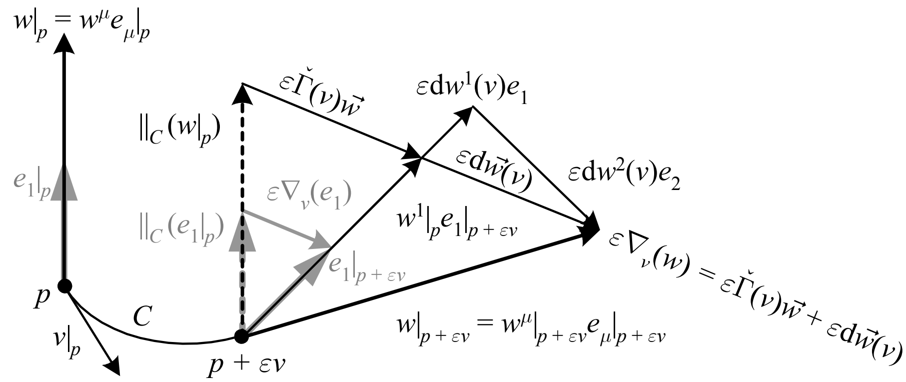

Figure 2.3: Relationships between the frame, parallel transport, covariant derivative, and connection for a vector w parallel to e_1 at a point p.

☼ The relation $\nabla_v w = \check{\Gamma}(v)\vec{w} + \mathrm{d}\vec{w}(v)$ can be viewed as roughly saying that the change in w under parallel transport is equal to the change in the frame relative to its parallel transport plus the change in the components of w in that frame.

If the 1-form ${\Gamma^\lambda}_\mu(v)$ itself is written using component notation, we arrive at the **connection coefficients**

$$ {\Gamma^\lambda}_{\mu\sigma} \equiv {\Gamma^\lambda}_\mu(e_\sigma) = \beta^\lambda(\nabla_{e_\sigma} e_\mu). \tag{2.9} $$

${\Gamma^\lambda}_{\mu\sigma}$ thus measures the λ^{th} component of the difference between e_μ and its parallel transport in the direction e_σ.

△ This notation is potentially confusing, as it makes ${\Gamma^\lambda}_{\mu\sigma}$ look like the components of a tensor, which it is not: it is a derivative of the component of the frame indexed by μ, and therefore is not only locally frame-dependent but also depends upon values of the frame at other points, so that it is not a multilinear mapping on its local arguments. Similarly, $\mathrm{d}\vec{w}$ looks like a frame-independent exterior derivative, but it is not: it is the exterior derivative of the frame-dependent components of w.

△ The ordering of the lower indices of ${\Gamma^\lambda}_{\mu\sigma}$ is not consistent across the literature (e.g. [9] vs [7]). This is sometimes not remarked upon, possibly due to the fact that in typical circumstances in general relativity (a coordinate frame and zero torsion, to be defined in Section 3.4), the connection coefficients are symmetric in their lower indices.

It is common to extend abstract index notation (see Section A.4) to be able to express the covariant derivative in terms of the connection coefficients as follows:

$$\nabla_{e_\mu} w = \mathrm{d}w^\gamma \left(e_\mu\right) e_\gamma + \Gamma^\nu{}_\gamma \left(e_\mu\right) w^\gamma e_\nu$$
$$\Rightarrow \nabla_a w^b \equiv \left(\nabla_{e_a} w\right)^b = e_a \left(w^b\right) + \Gamma^b{}_{ca} w^c \tag{2.10}$$
$$\Rightarrow \nabla_a w^b = \partial_a w^b + \Gamma^b{}_{ca} w^c$$

Here we have also defined $\partial_a f \equiv \partial_{e_a} f = \mathrm{d}f(e_a) = e_a(f)$. This notation is also sometimes supplemented to use a comma to indicate partial differentiation and a semicolon to indicate covariant differentiation, so that the above becomes

$$w^b{}_{;a} = w^b{}_{,a} + \Gamma^b{}_{ca} w^c. \tag{2.11}$$

The extension of index notation to derivatives has several potentially confusing aspects:

- ∇_a and ∂_a written alone are not 1-forms
- Greek indices indicate only that a specific basis (frame) has been chosen ([9] pp. 23-26), but do not distinguish between a general frame, where $\partial_\mu f \equiv \mathrm{d}f(e_\mu)$, and a coordinate frame, where $\partial_\mu f \equiv \partial f / \partial x^\mu$
- $\nabla_a w^b \equiv (\nabla_{e_a} w)^b$, so since $\nabla_v w$ is linear in v, $\nabla_a w^b$ is in fact a tensor of type $(1, 1)$; a more accurate notation might be $(\nabla w)^b{}_a$
- w^b in the expression $\partial_a w^b \equiv \mathrm{d}w^b(e_a)$ is not a vector, it is a set of frame-dependent component functions labeled by b whose change in the direction e_a is being measured
- The above means that, consistent with the definition of the connection coefficients, we have $\nabla_a e_b = 0 + e_c \Gamma^c{}_{ba}$, since the components of the frame itself by definition do not change
- As previously noted, neither $\Gamma^b{}_{ca}$ nor $\Gamma^b{}_{ca} w^c$ are tensors

We will nevertheless use this notation for many expressions going forward, as it is frequently used in general relativity.

△ It is important to remember that expressions involving ∇_a, ∂_a, and $\Gamma^c{}_{ba}$ must be handled carefully, as none of these are consistent with the original concept of indices denoting tensor components.

△ Some texts will distinguish between the labels of basis vectors and abstract index notation by using expressions such as $(e_i)^a$. We will not follow this practice, as it makes difficult the convenient method of matching indexes in expressions such as $\partial_a w^b \equiv \mathrm{d}w^b(e_a)$.

△ If we choose coordinates x^μ and use a coordinate frame so that $\partial_\mu \equiv \partial / \partial x^\mu$, we have the usual relation $\partial_\mu \partial_\nu f = \partial_\nu \partial_\mu f$. However, this is not necessarily implied by the Greek indices alone, which only indicate that a particular frame has been chosen. For index notation in general, mixed partials do not commute, since $\partial_a \partial_b f - \partial_b \partial_a f = e_a(e_b(f)) - e_b(e_a(f)) = [e_a, e_b](f) = [e_a, e_b]^c \partial_c f$, which only vanishes in a holonomic frame.

2.5 The parallel transporter in terms of the connection

We can also consider the parallel transport of a vector w along an infinitesimal curve C with tangent v. Referring to Fig. 2.3, we see that to order ε the components w^μ transform according to

$$\|^{\lambda}{}_{\mu}(C)\, w^{\mu} = w^{\lambda} - \varepsilon\Gamma^{\lambda}{}_{\mu}(v)\, w^{\mu}, \tag{2.12}$$

where v is tangent to the curve C, and these components are with respect to the frame at the new point after infinitesimal parallel transport. Using this relation, we can build up a frame-dependent expression for the parallel transporter for finite C by multiplying terms $(1 - \varepsilon\Gamma|_p)$ where $\Gamma|_p$ is used to denote the matrix $\Gamma^{\lambda}{}_{\mu}(v|_p)$ evaluated on the tangent $v|_p$ at successive points p along C. The limit of this process is the **path-ordered exponential**

$$\begin{aligned}\|^{\lambda}{}_{\mu}(C) &= \lim_{\varepsilon\to 0}(1-\varepsilon\Gamma|_{q-\varepsilon})(1-\varepsilon\Gamma|_{q-2\varepsilon})\cdots(1-\varepsilon\Gamma|_{p+\varepsilon})(1-\varepsilon\Gamma|_p)\\ &\equiv P\mathrm{exp}\left(-\int_C \Gamma^{\lambda}{}_{\mu}\right),\end{aligned} \tag{2.13}$$

whose definition is based on the expression for the exponential

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \lim_{\varepsilon\to 0}(1+\varepsilon x)^{1/\varepsilon}. \tag{2.14}$$

Note that the above expression for $\|^{\lambda}{}_{\mu}(C)$ exponentiates frame-dependent values in $gl(n,\mathbb{R})$ to yield a frame-dependent value in $GL(n,\mathbb{R})$.

2.6 Geodesics and normal coordinates

Following the example of the Lie derivative (see Section C.2), we can consider parallel transport of a vector v in the direction v as generating a local flow. More precisely, for any vector v at a point $p \in M$, there is a curve $\phi_v(t)$, unique for some $-\varepsilon < t < \varepsilon$, such that $\phi_v(0) = p$ and $\dot{\phi}_v(t) = \|_\phi(v)$, the last expression indicating that the tangent to ϕ_v at t is equal to the parallel transport of v along ϕ_v from $\phi_v(0)$ to $\phi_v(t)$. This curve is called a **geodesic**, and its tangent vectors are all parallel transports of each other. This means that for all tangent vectors v to the curve, $\nabla_v v = 0$, so that geodesics are "the closest thing to straight lines" on a manifold with parallel transport.

Now we can define the **exponential map** at p to be $\exp(v) \equiv \phi_v(1)$, which will be well-defined for values of v around the origin that map to some $U \subset M$ containing p. Finally, choosing a basis for T_pU provides an isomorphism $T_pU \cong \mathbb{R}^n$, allowing us to define **geodesic normal coordinates** (AKA normal coordinates) $\exp^{-1}\colon U \to \mathbb{R}^n$. It can be shown (see [6] Vol. 1 pp148-149) that in a coordinate frame at the origin p of geodesic normal coordinates, we have $\Gamma^{\lambda}{}_{\mu\sigma} = -\Gamma^{\lambda}{}_{\sigma\mu}$; this implies that for zero torsion (to be defined in Section 3.4), the connection coefficients vanish at p.

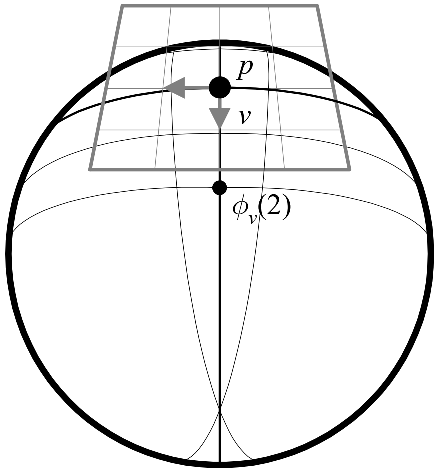

FIGURE 2.4: Geodesic normal coordinates at p map points on a manifold to vectors at p tangent to the geodesic passing through both points. In the figure $\exp(2v) = \phi_v(2)$, so the coordinate of the point $\phi_v(2) \in M$ is $2v \in T_pM$.

2.7 Summary

In general, a "manifold with connection" is one with an additional structure that "connects" the different tangent spaces of the manifold to one another in a linear fashion. Specifying any one of the above connection quantities, the covariant derivative, or the parallel transporter equivalently determines this structure. The following tables summarize the situation.

Construct	Argument(s)	Value	Dependencies
$\|_C$	$v \in T_pM$	$\|_C (v) \in T_qM$	Path C from p to q
$\|^{\lambda}{}_{\mu}$	Path C	$\|^{\lambda}{}_{\mu} (C) \in GL$	Frame on M
∇_v	$w \in TM$	$\nabla_v w \in T_pM$	$v \in T_pM$
∇	$v \in T_pM$, $w \in TM$	$\nabla_v w \in T_pM$	None
$\Gamma^{\lambda}{}_{\mu}$	$v \in T_pM$	$\Gamma^{\lambda}{}_{\mu} (v) \in gl$	Frame on M
$\check{\Gamma} (v)$	$\vec{w} \in T_pM$	$\check{\Gamma} (v) \vec{w} \in T_pM$	Frame on M, $v \in T_pM$
$\Gamma^{\lambda}{}_{\mu\sigma}$	None	Connection coefficient	Frame on M

TABLE 2.1: Constructions related to the connection. Each construct above is considered at a point p; to determine a manifold with connection it must be defined for every point in M.

Below we review the intuitive meanings of the various vector derivatives.

Vector derivative	Meaning
$L_v w \equiv \lim_{\varepsilon \to 0} \left(w \left\|_{p+\varepsilon v} - \mathrm{d}\Phi_\varepsilon \left(w \right\|_p \right) \right) / \varepsilon$	The difference between w and its transport by the local flow of v.
$\nabla_v w \equiv \lim_{\varepsilon \to 0} \left(w \left\|_{p+\varepsilon v} - \parallel_C \left(w \right\|_p \right) \right) / \varepsilon$	The difference between w and its parallel transport in the direction v.
$\frac{\mathrm{D}}{\mathrm{d}t} w \equiv \mathrm{D}_t w \equiv \nabla_{\dot{C}(t)} w$	The difference between w and its parallel transport in the direction tangent to $C(t)$.
${\Gamma^\lambda}_\mu (v) \equiv \beta^\lambda (\nabla_v e_\mu)$	The λ^{th} component of the difference between e_μ and its parallel transport in the direction v.
$\check{\Gamma}(v) \equiv \nabla_v (T_p M)$	The infinitesimal linear transformation on the tangent space that takes the parallel transported frame to the frame in the direction v.
$\check{\Gamma}(v) \vec{w} \equiv {\Gamma^\lambda}_\mu (v) w^\mu e_\lambda = (\nabla_v e_\mu) w^\mu$	The difference between the frame and its parallel transport in the direction v, weighted by the components of w.
${\Gamma^\lambda}_{\mu\sigma} \equiv {\Gamma^\lambda}_\mu (e_\sigma) = \beta^\lambda (\nabla_\sigma e_\mu)$	The λ^{th} component of the difference between e_μ and its parallel transport in the direction e_σ.
$\mathrm{d}\vec{w}(v) \equiv \mathrm{d}w^\mu (v) e_\mu$	The change in the frame-dependent components of w in the direction v.
$\partial_a w^b \equiv \mathrm{d}w^b(e_a)$	The change in the b^{th} frame-dependent component of w in the direction e_a.
$\nabla_a w^b \equiv (\nabla_{e_a} w)^b$	The b^{th} component of the difference between w and its parallel transport in the direction e_a.

TABLE 2.2: Definitions and meanings of vector derivatives.

Other quantities in terms of the connection:

- $\nabla_v w = \mathrm{d}\vec{w}(v) + \check{\Gamma}(v) \vec{w}$
- $\nabla_a w^b = \partial_a w^b + {\Gamma^b}_{ca} w^c$
- ${\parallel^\lambda}_\mu (C) w^\mu = w^\mu - \varepsilon {\Gamma^\lambda}_\mu (v) w^\mu$ (for infinitesimal C with tangent v)
- ${\parallel^\lambda}_\mu (C) w^\mu = P\exp\left(-\int_C {\Gamma^\lambda}_\mu \right) w^\mu$

3 Manifolds with connection

All of the above constructs used to define a manifold with connection manipulate vectors, which means they can be naturally extended to operate on arbitrary tensor fields on M. This is the usual approach taken in general relativity; however, one can alternatively focus on k-forms on M, an approach that generalizes more directly to gauge theories in physics. This viewpoint is sometimes called the **Cartan formalism**. We will cover both approaches.

△ Note that a manifold with connection includes no concept of length or distance (a metric). It is important to remember that unless noted, nothing in this section depends upon this extra structure.

3.1 The covariant derivative on the tensor algebra

If we define the covariant derivative of a function to coincide with the normal derivative, i.e. $\nabla_a f \equiv \partial_a f$, then we can use the Leibniz rule to define the covariant derivative of a 1-form. This is sometimes described as making the covariant derivative "commute with contractions," where for a 1-form φ and a vector v we require

$$\begin{aligned}\nabla_a\left(\varphi_b v^b\right) &\equiv \left(\nabla_a \varphi_b\right) v^b + \varphi_b\left(\nabla_a v^b\right)\\ &= \left(\nabla_a \varphi_b\right) v^b + \varphi_b\left(\partial_a v^b + \Gamma^b{}_{ca} v^c\right).\end{aligned} \tag{3.1}$$

At the same time, choosing a frame and treating φ_b and v^b as frame-dependent functions on M, we have

$$\begin{aligned}\nabla_a\left(\varphi_b v^b\right) &\equiv \partial_a\left(\varphi_b v^b\right)\\ &= \left(\partial_a \varphi_b\right) v^b + \varphi_b\left(\partial_a v^b\right),\end{aligned} \tag{3.2}$$

so that equating the two we arrive at

$$\nabla_a \varphi_b \equiv \partial_a \varphi_b - \Gamma^c{}_{ba} \varphi_c. \tag{3.3}$$

As with vectors, the partial derivative $\partial_a \varphi_b$ acts upon the frame-dependent components of the 1-form.

We can then extend the covariant derivative to be a derivation on the tensor algebra (see Section C.1) by following the above logic for each covariant and contravariant component:

$$\begin{aligned}\nabla_a T^{b_1 \ldots b_m}{}_{c_1 \ldots c_n} \equiv\; &\partial_a T^{b_1 \ldots b_m}{}_{c_1 \ldots c_n}\\ &+ \sum_{j=1}^{m} \Gamma^{b_j}{}_{da} T^{b_1 \ldots b_{j-1} d b_{j+1} \ldots b_m}{}_{c_1 \ldots c_n}\\ &- \sum_{j=1}^{n} \Gamma^{d}{}_{c_j a} T^{b_1 \ldots b_m}{}_{c_1 \ldots c_{j-1} d c_{j+1} \ldots c_n}\end{aligned} \tag{3.4}$$

Note that since the covariant derivative of a 0-form is $\nabla_a f = \partial_a f = \partial_{e_a} f = e_a(f)$, we then have $\nabla_v f = v^a \nabla_a f = v^a e_a(f) = v(f)$.

The concept of parallel transport along a curve C can be extended to the tensor algebra as well, by parallel transporting all vector arguments backwards to the starting point of C, applying the tensor, then parallel transporting the resulting vectors forward to the endpoint of C. So for example the parallel transport of a tensor $T^a{}_b$ is defined as

$$\begin{aligned}\|_C\left(T^a{}_b\right) &\equiv \|^a{}_c\,(C)\, T^c{}_d\, \|^d{}_b\,(-C)\\ &= \left(1 - \varepsilon \Gamma^a{}_c(v)\right) T^c{}_d \left(1 + \varepsilon \Gamma^d{}_b(v)\right),\end{aligned} \tag{3.5}$$

where for infinitesimal C with tangent v we have $\|_C^{-1} = \|_{-C} = 1 + \varepsilon \check{\Gamma}(v)$ since $\|_C = 1 - \varepsilon \check{\Gamma}(v)$.

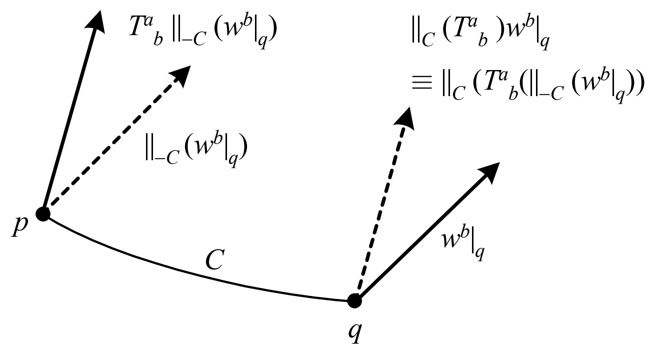

FIGURE 3.1: The parallel transport of a tensor can be defined by parallel transporting all vector arguments backwards to the starting point, applying the tensor, then parallel transporting the resulting vectors forward to the endpoint.

With this definition, the covariant derivative $\nabla_a T$ can be viewed as "the difference between T and its parallel transport in the direction e_a."

△ It can sometimes be confusing when using the extended covariant derivative as to what type of tensor it is being applied to. For example, w^b in the expression $\partial_a w^b$ is not a vector, it is a set of frame-dependent functions labeled by b; yet this expression can in theory also be written $\nabla_a w^b$, in which case there is no indication that the covariant derivative is acting on these functions instead of the vector w^b.

△ When the covariant derivative is used as a derivation on the tensor algebra, care must be taken with relations, since their forms can change considerably based upon what arguments are applied and whether index notation is used. In particular, $(\nabla_a \nabla_b - \nabla_b \nabla_a) f = \nabla_a(\partial_b f) - \nabla_b(\partial_a f)$ is not a "mixed partials" expression, since $(\partial_a f)$ is a 1-form. And as we will see, $(\nabla_a \nabla_b - \nabla_b \nabla_a) f$ is a different construction than $(\nabla_a \nabla_b - \nabla_b \nabla_a) w^c$, which is different from $(\nabla_u \nabla_v - \nabla_v \nabla_u) w$. It is important to realize that an expression such as $\nabla_a \nabla_b - \nabla_b \nabla_a$ without context has no unambiguous meaning.

△ It is important to remember that since expressions like $\partial_a w^b$ and $\Gamma^c{}_{ba}$ are not tensors, applying ∇_d to them is not well-defined (unless we consider them as arrays of functions and are applying $\nabla_d = \partial_d$).

3.2 The exterior covariant derivative of vector-valued forms

A vector field w on M can be viewed as a vector-valued 0-form. As noted previously, the covariant derivative $\nabla_v w$ is linear in v and depends only on its local value, and so can be viewed as a vector-valued 1-form $\mathrm{D}\vec{w}(v) \equiv \nabla_v w$. $\mathrm{D}\vec{w}$ is called the **exterior covariant derivative** of the vector-valued 0-form $\vec{w}$. This definition is then extended to vector-valued k-forms $\vec{\varphi}$ by following the example of the exterior derivative d (see Section C.5):

$$
\begin{aligned}
&\mathrm{D}\vec{\varphi}\left(v_0,\ldots,v_k\right)\\
&\equiv \sum_{j=0}^{k}(-1)^j\,\nabla_{v_j}\left(\vec{\varphi}\left(v_0,\ldots,v_{j-1},v_{j+1},\ldots,v_k\right)\right)\\
&\quad+\sum_{i<j}(-1)^{i+j}\,\vec{\varphi}\left(\left[v_i,v_j\right],v_0,\ldots,v_{i-1},v_{i+1},\ldots,v_{j-1},v_{j+1},\ldots,v_k\right)
\end{aligned}
\tag{3.6}
$$

For example, if $\vec{\varphi}$ is a vector-valued 1-form,

$$
\mathrm{D}\vec{\varphi}\left(v,w\right) \equiv \nabla_v\vec{\varphi}\left(w\right) - \nabla_w\vec{\varphi}\left(v\right) - \vec{\varphi}\left(\left[v,w\right]\right). \tag{3.7}
$$

So while the first term of $\mathrm{d}\varphi$ takes the difference between the scalar values of $\varphi(w)$ along v, the first term of $\mathrm{D}\vec{\varphi}$ takes the difference between the vector values of $\vec{\varphi}(w)$ along v after parallel transporting them to the same point (which is required to compare them). At a point p, $\mathrm{D}\vec{\varphi}\left(v,w\right)$ can thus be viewed as the "sum of $\vec{\varphi}$ on the boundary of the surface defined by its arguments after being parallel transported back to p," and if we use $\|_{\varepsilon v}$ to denote parallel transport along an infinitesimal curve with tangent v, we can write

$$
\begin{aligned}
\varepsilon^2\mathrm{D}\vec{\varphi}\left(v,w\right) = &\,\|_{-\varepsilon v}\vec{\varphi}\left(\varepsilon w\left|_{p+\varepsilon v}\right.\right) - \vec{\varphi}\left(\varepsilon w\left|_{p}\right.\right)\\
&- \|_{-\varepsilon w}\vec{\varphi}\left(\varepsilon v\left|_{p+\varepsilon w}\right.\right) + \vec{\varphi}\left(\varepsilon v\left|_{p}\right.\right)\\
&- \vec{\varphi}\left(\varepsilon^2\left[v,w\right]\right).
\end{aligned}
\tag{3.8}
$$

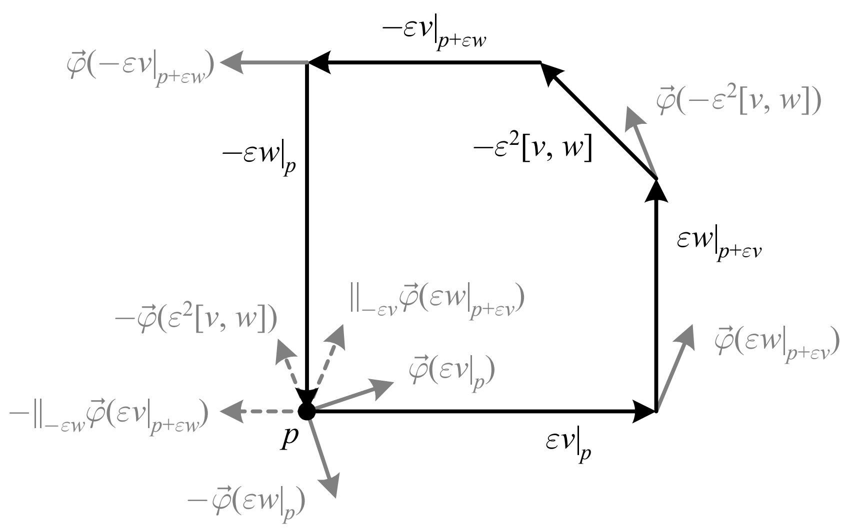

FIGURE 3.2: The exterior covariant derivative $\mathrm{D}\vec{\varphi}\left(v,w\right)$ sums the vectors $\vec{\varphi}$ along the boundary of the surface defined by v and w by parallel transporting them to the same point. Note that the "completion of the parallelogram" $[v,w]$ is already of order ε^2, so its parallel transport has no effect to this order.

From its definition, it is clear that $\mathrm{D}\vec{\varphi}$ is a frame-independent quantity. In terms of the connection, we must consider $\vec{w}$ as a frame-dependent $\mathbb{R}^n$-valued 0-form, so that

$$
\mathrm{D}\vec{w}\left(v\right) = \nabla_v w = \mathrm{d}\vec{w}\left(v\right) + \check{\Gamma}\left(v\right)\vec{w}. \tag{3.9}
$$

For a $\mathbb{R}^n$-valued k-form $\vec{\varphi}$ we find that

$$\mathrm{D}\vec{\varphi} = \mathrm{d}\vec{\varphi} + \check{\Gamma} \wedge \vec{\varphi}, \tag{3.10}$$

where the exterior derivative is defined to apply to the frame-dependent components, i.e. $\mathrm{d}\vec{\varphi}(v_0 \dots v_k) \equiv \mathrm{d}\varphi^\mu(v_0 \dots v_k) e_\mu$. Recall that $\check{\Gamma}$ is a $gl(n,\mathbb{R})$-valued 1-form, so that for example if $\vec{\varphi}$ is a $\mathbb{R}^n$-valued 1-form then

$$\begin{aligned} (\check{\Gamma} \wedge \vec{\varphi})(v,w) &\equiv \check{\Gamma}(v)\, \vec{\varphi}(w) - \check{\Gamma}(w)\, \vec{\varphi}(v) \\ &= \Gamma^\lambda{}_\mu(v)\, \varphi^\mu(w) - \Gamma^\lambda{}_\mu(w)\, \varphi^\mu(v)\,. \end{aligned} \tag{3.11}$$

△ As with the covariant derivative, it is important to remember that $\mathrm{D}\vec{\varphi}$ is frame-independent while $\mathrm{d}\vec{\varphi}$ and $\check{\Gamma}$ are not.

The set of vector-valued forms can be viewed as an infinite-dimensional algebra by defining multiplication via the vector field commutator; it turns out that D does not satisfy the Leibniz rule in this algebra and so is not a derivation (see Appendix C.1). However, following the above reasoning one can extend the definition of D to the algebra of tensor-valued forms, or the subset of anti-symmetric tensor-valued forms; D then is a derivation with respect to the tensor product in the former case and a graded derivation with respect to the exterior product in the latter case. We will not pursue either of these two generalizations.

3.3 The exterior covariant derivative of algebra-valued forms

Recalling from Section 3.1 the definition of parallel transport of a tensor, we can view a $gl(n,\mathbb{R})$-valued 0-form $\check{\Theta}$ as a tensor of type $(1,1)$, so that the infinitesimal parallel transport of $\check{\Theta}$ along C with tangent v is

$$\|_C\,(\check{\Theta}) = \left(1 - \varepsilon\check{\Gamma}(v)\right) \check{\Theta} \left(1 + \varepsilon\check{\Gamma}(v)\right). \tag{3.12}$$

We can now follow the reasoning used to define the covariant derivative of a vector in terms of the connection

$$\begin{aligned} \nabla_v w &\equiv \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(w\,|_{p+\varepsilon v} - \|_C\; w\,|_p \right) \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(\vec{w}\,|_{p+\varepsilon v} - \left(1 - \varepsilon\check{\Gamma}(v)\right) \vec{w}\,|_p \right) \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(w^\mu\,|_{p+\varepsilon v} - w^\mu\,|_p + \varepsilon \Gamma^\mu{}_\lambda(v)\, w^\lambda\,|_p \right) e_\mu\,|_{p+\varepsilon v} \\ &= \mathrm{d}\vec{w}(v) + \check{\Gamma}(v)\, \vec{w} \end{aligned} \tag{3.13}$$

to give the covariant derivative of a $gl(n,\mathbb{R})$-valued 0-form

$$\begin{aligned} \nabla_v \check{\Theta} &\equiv \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(\check{\Theta}\,|_{p+\varepsilon v} - \|_C\,\left(\check{\Theta}\,|_p\right) \right) \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(\check{\Theta}\,|_{p+\varepsilon v} - \left(1 - \varepsilon\check{\Gamma}(v)\right) \check{\Theta}\,|_p \left(1 + \varepsilon\check{\Gamma}(v)\right) \right) \\ &= \mathrm{d}\check{\Theta}(v) + \check{\Gamma}(v)\, \check{\Theta} - \check{\Theta}\check{\Gamma}(v) \\ &= \mathrm{d}\check{\Theta}(v) + \left[\check{\Gamma}, \check{\Theta}\right](v) \\ &= \mathrm{d}\check{\Theta}(v) + \left(\check{\Gamma}[\wedge]\check{\Theta}\right)(v)\,. \end{aligned} \tag{3.14}$$

Here we have only kept terms to order ε, followed previous convention to define $\mathrm{d}\check{\Theta}(v) \equiv \mathrm{d}\Theta^{\mu}{}_{\lambda}\beta^{\lambda}e_{\mu}$, and defined the Lie commutator $[\check{\Gamma}, \check{\Theta}]$ in terms of the multiplication of the $gl(n,\mathbb{R})$-valued forms $\check{\Gamma}$ and $\check{\Theta}$, which (see Section A.9 for notation) as a 1-form is equivalent to $\check{\Gamma}[\wedge]\check{\Theta}$. $\nabla_v\check{\Theta}$ is then "the difference between the linear transformation $\check{\Theta}$ and its parallel transport in the direction v."

The above definition of the covariant derivative can then be extended to arbitrary $gl(n,\mathbb{R})$-valued k-forms by defining

$$\mathrm{D}\check{\Theta} \equiv \mathrm{d}\check{\Theta} + \check{\Gamma}[\wedge]\check{\Theta}, \tag{3.15}$$

which can be shown to be equivalent to the construction used for $\mathbb{R}^n$-valued k-forms in Section 3.2. For example, for a $gl(n,\mathbb{R})$-valued 1-form $\check{\Theta}$, we have

$$\mathrm{D}\check{\Theta}(v,w) \equiv \nabla_v\check{\Theta}(w) - \nabla_w\check{\Theta}(v) - \check{\Theta}([v,w]), \tag{3.16}$$

with the covariant derivatives acting on the value of $\check{\Theta}$ as a tensor of type $(1,1)$. So at a point p, $\mathrm{D}\check{\Theta}(v,w)$ can be viewed as the "sum of $\check{\Theta}$ on the boundary of the surface defined by its arguments after being parallel transported back to p." With respect to the set of $gl(n,\mathbb{R})$-valued forms under the exterior product using the Lie commutator $[\wedge]$, D is a graded derivation and for a $gl(n,\mathbb{R})$-valued k-form $\check{\Theta}$ satisfies the Leibniz rule

$$\mathrm{D}(\check{\Theta}[\wedge]\check{\Psi}) = \mathrm{D}\check{\Theta}[\wedge]\check{\Psi} + (-1)^k\,\check{\Theta}[\wedge]\mathrm{D}\check{\Psi}. \tag{3.17}$$

3.4 Torsion

Given a frame e_μ, we can view the dual frame β^μ as a vector-valued 1-form that simply returns its vector argument:

$$\vec{\beta}(v) \equiv \beta^\mu(v)\,e_\mu = v. \tag{3.18}$$

Clearly this is a frame-independent object. The **torsion** is then defined to be the exterior covariant derivative

$$\vec{T} \equiv \mathrm{D}\vec{\beta}. \tag{3.19}$$

In terms of the connection, we must consider $\vec{\beta}$ as a frame-dependent $\mathbb{R}^n$-valued 1-form, which gives us the torsion as a $\mathbb{R}^n$-valued 2-form

$$\vec{T} = \mathrm{d}\vec{\beta} + \check{\Gamma} \wedge \vec{\beta}. \tag{3.20}$$

This definition of $\vec{T}$ is sometimes called **Cartan's first structure equation**.

In terms of the covariant derivative, the torsion 2-form is

$$\begin{aligned}\vec{T}(v,w) &\equiv \nabla_v\left(\vec{\beta}(w)\right) - \nabla_w\left(\vec{\beta}(v)\right) - \vec{\beta}([v,w]) \\ &= \nabla_v w - \nabla_w v - [v,w]\,.\end{aligned} \tag{3.21}$$

For a torsion-free connection in a holonomic frame, we then have $\nabla_\sigma e_\mu = \nabla_\mu e_\sigma$, which means that the connection coefficients are symmetric in their lower indices, i.e.

$$\Gamma^{\lambda}{}_{\mu\sigma} \equiv \beta^\lambda(\nabla_\sigma e_\mu) = \beta^\lambda(\nabla_\mu e_\sigma) = \Gamma^{\lambda}{}_{\sigma\mu}. \tag{3.22}$$

For this reason, a torsion-free connection is also called a **symmetric connection**.

From the definition in terms of the exterior covariant derivative, we can view the torsion as the "sum of the boundary vectors of the surface defined by its arguments after being parallel transported back to p," i.e. the torsion measures the amount by which the boundary of a loop fails to close after being parallel transported. From the definition in terms of the covariant derivative, we arrive in the figure below at another interpretation where, like the Lie derivative $L_v w$ (see Section C.2), $\vec{T}(v,w)$ "completes the parallelogram" formed by its vector arguments, but this parallelogram is formed by parallel transport instead of local flow. Note however that the torsion vector has the opposite sign as the Lie derivative.

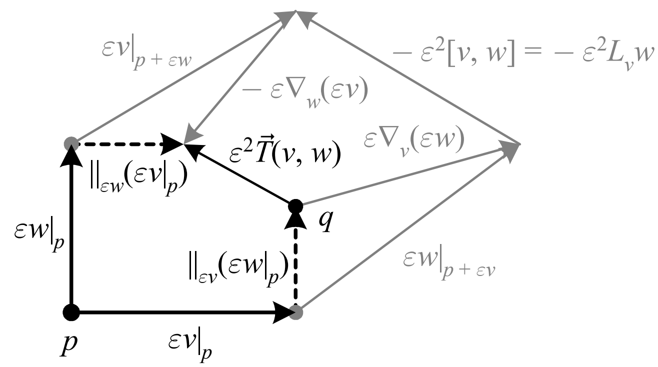

FIGURE 3.3: The torsion vector $\vec{T}(v,w)$, constructed above starting at the point q, "completes the parallelogram" formed by parallel transport. $\|_{\varepsilon v}$ denotes parallel transport along an infinitesimal curve with tangent v.

Zero torsion then means that moving infinitesimally along v followed by the parallel transport of w is the same as moving infinitesimally along w followed by the parallel transport of v. Non-zero torsion signifies that "a loop made of parallel transported vectors is not closed."

As this geometric interpretation suggests, and as is evident from the expression $\vec{T} \equiv \mathrm{D}\vec{\beta}$, one can verify algebraically that despite being defined in terms of derivatives $\vec{T}(v,w)$ in fact only depends on the local values of v and w, and thus can be viewed as a tensor of type $(1,2)$:

$$T^c{}_{ab} v^a w^b \equiv v^a \nabla_a w^c - w^a \nabla_a v^c - [v,w]^c \tag{3.23}$$

Another relation can be obtained for the torsion tensor by applying its vector value to a function f before moving into index notation:

$$\begin{aligned} \vec{T}(v,w)(f) &\equiv (\nabla_v w)(f) - (\nabla_w v)(f) - [v,w](f) \\ \Rightarrow T^c{}_{ab} v^a w^b \nabla_c f &= \left(v^a \nabla_a w^b\right) \nabla_b f - \left(w^b \nabla_b v^a\right) \nabla_a f \\ &\quad - \left[v^a \nabla_a \left(w^b \nabla_b f\right) - w^b \nabla_b \left(v^a \nabla_a f\right)\right] \\ \Rightarrow T^c{}_{ab} \nabla_c f &= \nabla_b \nabla_a f - \nabla_a \nabla_b f \end{aligned} \tag{3.24}$$

Here we have used the Leibniz rule and recalled that $v(f) = \nabla_v f = v^a \nabla_a f$ and $[v,w](f) = v(w(f)) - w(v(f))$ (see Section B.2). In terms of the connection coefficients $\Gamma^c{}_{ab} = \beta^c \nabla_b e_a$ we have

$$
\begin{aligned}
T^c{}_{ab} &= \beta^c \vec{T}\,(e_a, e_b) \\
&= \beta^c \nabla_a e_b - \beta^c \nabla_b e_a - \beta^c [e_a, e_b] \\
&= \Gamma^c{}_{ba} - \Gamma^c{}_{ab} - [e_a, e_b]^c .
\end{aligned}
\tag{3.25}
$$

△ Note that zero torsion thus always means that $\nabla_a \nabla_b f = \nabla_b \nabla_a f$ (and $[v, w] = L_v w = \nabla_v w - \nabla_w v$), but it only means $\Gamma^\lambda{}_{\mu\sigma} = \Gamma^\lambda{}_{\sigma\mu}$ in a holonomic frame.

In the above figure, the failure of the parallel transported vectors to meet can be viewed as either due to their lengths changing or due to their being rotated out of the plane of the figure. As we will see, the latter interpretation is more relevant for Riemannian manifolds, where parallel transport leaves lengths invariant. In Einstein-Cartan theory in physics, non-zero torsion is associated with spin in matter. A suggestive example along these lines that highlights the rotation aspect of torsion is Euclidean $\mathbb{R}^3$ with parallel transport defined by translation, except in the x direction where parallel transport rotates a vector clockwise by an angle proportional to the distance transported. As we will see in the next section, this parallel transport has torsion but no curvature.

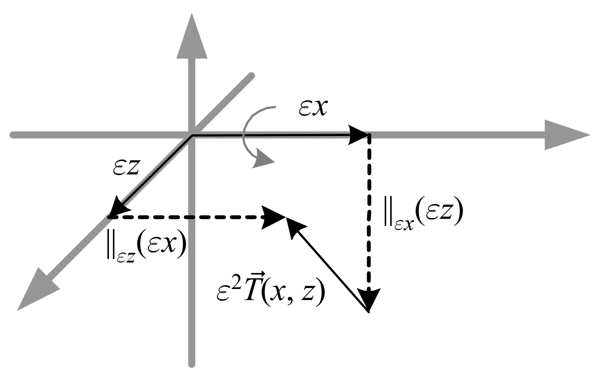

FIGURE 3.4: An example of non-zero torsion suggestive of spin.

Zero torsion means that $L_v w = [v, w] = \nabla_v w - \nabla_w v$ due to the symmetric connection coefficients canceling. This extends to the Lie derivative of a general tensor, so that in the case of zero torsion we have

$$
\begin{aligned}
L_v T^{a_1 \dots a_m}{}_{b_1 \dots b_n} &= v^c \nabla_c T^{a_1 \dots a_m}{}_{b_1 \dots b_n} \\
&\quad - \sum_{j=1}^{m} \left(\nabla_c v^{a_j}\right) T^{a_1 \dots a_{j-1} c a_{j+1} \dots a_m}{}_{b_1 \dots b_n} \\
&\quad + \sum_{j=1}^{n} \left(\nabla_{b_j} v^c\right) T^{a_1 \dots a_m}{}_{b_1 \dots b_{j-1} c b_{j+1} \dots b_n} .
\end{aligned}
\tag{3.26}
$$

3.5 Curvature

The exterior covariant derivative D parallel transports its values on the boundary before summing them, and therefore we do not expect it to mimic the property $\mathrm{d}^2 = 0$ (see Section C.4).

Indeed it does not; instead, for a vector field w viewed as a vector-valued 0-form $\vec{w}$, we have

$$\left(\mathrm{D}^2\vec{w}\right)(u,v) \equiv \check{R}\,(u,v)\,\vec{w} = \nabla_u\nabla_v w - \nabla_v\nabla_u w - \nabla_{[u,v]}w, \tag{3.27}$$

which defines the **curvature 2-form** $\check{R}$, which is $gl(\mathbb{R}^n)$-valued. From its definition, $\check{R}\vec{w}$ is a frame-independent quantity, and thus if $\vec{w}$ is considered as a vector-valued 0-form, $\check{R}$ is frame-independent as well. In the (more common) case that we view $\vec{w}$ as a frame-dependent $\mathbb{R}^n$-valued 0-form, $\check{R}$ must be considered to be $gl(n,\mathbb{R})$-valued, and is thus a frame-dependent matrix. A connection with zero curvature is called **flat**, as is any region of M with a flat connection.

For a general $\mathbb{R}^n$-valued form $\vec{\varphi}$ it is not hard to arrive at an expression for $\check{R}$ in terms of the connection:

$$\mathrm{D}^2\vec{\varphi} = \left(\mathrm{d}\check{\Gamma} + \check{\Gamma}\wedge\check{\Gamma}\right)\wedge\vec{\varphi} \equiv \check{R}\wedge\vec{\varphi} \tag{3.28}$$

Note that $\mathrm{D}\check{\Gamma} = \mathrm{d}\check{\Gamma} + \check{\Gamma}[\wedge]\check{\Gamma}$ is a similar but distinct construction, since e.g.

$$(\check{\Gamma}\wedge\check{\Gamma})\,(v,w) = \check{\Gamma}\,(v)\,\check{\Gamma}\,(w) - \check{\Gamma}\,(w)\,\check{\Gamma}\,(v)\,, \tag{3.29}$$

while

$$\begin{aligned}(\check{\Gamma}[\wedge]\check{\Gamma})\,(v,w) &= [\check{\Gamma}\,(v)\,,\check{\Gamma}\,(w)] - [\check{\Gamma}\,(w)\,,\check{\Gamma}\,(v)]\\ &= 2(\check{\Gamma}\wedge\check{\Gamma})\,(v,w)\,.\end{aligned} \tag{3.30}$$

Thus we have

$$\begin{aligned}\check{R} &\equiv \mathrm{d}\check{\Gamma} + \check{\Gamma}\wedge\check{\Gamma}\\ &= \mathrm{d}\check{\Gamma} + \frac{1}{2}\check{\Gamma}[\wedge]\check{\Gamma}.\end{aligned} \tag{3.31}$$

The definition of $\check{R}$ in terms of $\check{\Gamma}$ is sometimes called **Cartan's second structure equation**. An immediate property from the definition of $\check{R}$ is

$$\check{R}(u,v) = -\check{R}(v,u), \tag{3.32}$$

which allows us to write e.g. for a vector-valued 1-form $\vec{\varphi}$

$$\begin{aligned}\left(\mathrm{D}^2\vec{\varphi}\right)(u,v,w) &\equiv \left(\check{R}\wedge\vec{\varphi}\right)(u,v,w)\\ &= \check{R}\,(u,v)\,\vec{\varphi}(w) + \check{R}\,(v,w)\,\vec{\varphi}(u) + \check{R}\,(w,u)\,\vec{\varphi}(v).\end{aligned} \tag{3.33}$$

Constructing the same picture as can be done for the double exterior derivative (see Section C.4), we put

$$\mathrm{D}^2\vec{w} \equiv \mathrm{D}\vec{\varphi},$$

where

$$\vec{\varphi}(v) \equiv \mathrm{D}\vec{w}(v) = \nabla_v w.$$

Expanding both derivatives in terms of parallel transport, we find in the following figure that as we sum values around the boundary of the surface defined by its arguments, D^2 fails to cancel the endpoint and starting point at the far corner. Examining the values of these non-canceling points, we can view the curvature as "the difference between w when parallel transported around the two opposite edges of the boundary of the surface defined by its arguments."

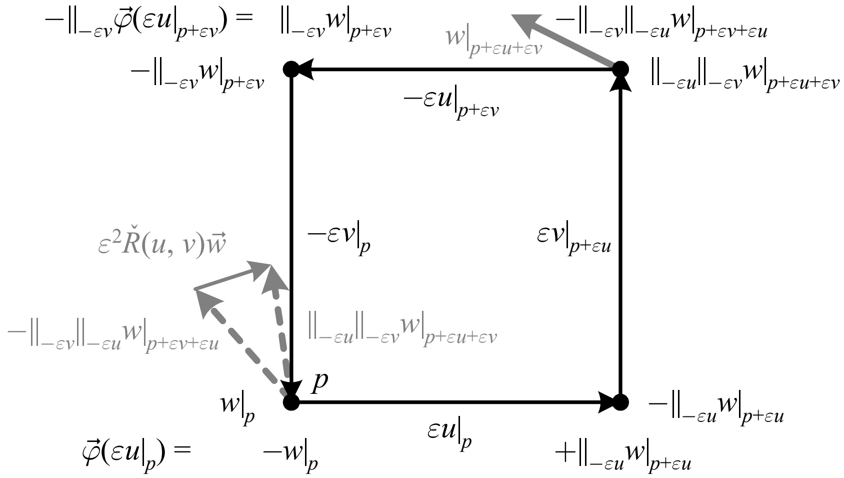

FIGURE 3.5: $\check{R}(u,v)\,\vec{w} = \left(\mathrm{D}^2\vec{w}\right)(u,v)$ is "the difference between w when parallel transported around the two opposite edges of the boundary of the surface defined by its arguments." In the figure we assume vanishing Lie bracket for simplicity, so that $v\,|_{p+\varepsilon u+\varepsilon v} = v\,|_{p+\varepsilon v+\varepsilon u}$.

In terms of the connection, we can use the path integral formulation to examine the parallel transporter around the closed path $L \equiv \partial S$ defined by the surface $S \equiv (\varepsilon u \wedge \varepsilon v)$ to order ε^2. This calculation after some work (see [4] pp. 51-53) yields

$$\begin{aligned} \|_L\,(w) &= P\mathrm{exp}\left(-\int_L \check{\Gamma}\right)\vec{w} \\ &= w - \int_S \left(\mathrm{d}\check{\Gamma} + \check{\Gamma}\wedge\check{\Gamma}\right)\vec{w} \\ &= w - \varepsilon^2\check{R}\,(u,v)\,\vec{w}, \end{aligned} \tag{3.34}$$

where we have dropped the indices since L is a closed path and thus $\|_L$ is basis-independent. Thus the curvature can be viewed as "the difference between w and its parallel transport around the boundary of the surface defined by its arguments."

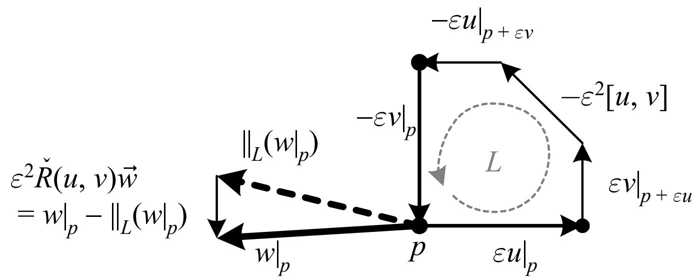

FIGURE 3.6: $\check{R}\,(u,v)\,\vec{w}$ is "the difference between w and its parallel transport around the boundary of the surface defined by its arguments."

As this picture suggests, one can verify algebraically that the value of $\check{R}\,(u,v)\,\vec{w}$ at a point p only depends upon the value of w at p, even though it can be defined in terms of ∇w, which depends upon nearby values of w. Similarly, $\check{R}\,(u,v)\,\vec{w}$ at a point p only depends upon the values of u and v at p, even though it can be defined in terms of $[u,v]$, which depends upon their vector field values (note that $\nabla_u \nabla_v w$ depends upon the vector field values of both v and w). Finally, $\check{R}$ (as a $gl(\mathbb{R}^n)$-valued 2-form) is frame-independent, even though it can be defined in terms of $\check{\Gamma}$, which is not. Thus the curvature can be viewed as a tensor of type $(1,3)$, called the **Riemann curvature tensor** (AKA Riemann tensor, curvature tensor, Riemann–Christoffel tensor):

$$\begin{aligned} R^c{}_{dab} u^a v^b w^d &\equiv u^a \nabla_a \left(v^b \nabla_b w^c \right) - v^b \nabla_b \left(u^a \nabla_a w^c \right) - [u,v]^d \nabla_d w^c \\ &= u^a v^b \nabla_a \nabla_b w^c - u^a v^b \nabla_b \nabla_a w^c + T^d{}_{ab} u^a v^b \nabla_d w^c \\ \Rightarrow R^c{}_{dab} w^d &= \left(\nabla_a \nabla_b - \nabla_b \nabla_a + T^d{}_{ab} \nabla_d \right) w^c \end{aligned} \tag{3.35}$$

Here we have used the Leibniz rule and recalled that $[u,v]^d = u^a \nabla_a v^d - v^b \nabla_b u^d - T^d{}_{ab} u^a v^b$.

To obtain an expression in terms of the connection coefficients, we first examine the double covariant derivative, recalling that $\nabla_b w^c$ is a tensor:

$$\begin{aligned} \nabla_a \left(\nabla_b w^c \right) &= \partial_a \nabla_b w^c + \Gamma^c{}_{fa} \nabla_b w^f - \Gamma^f{}_{ba} \nabla_f w^c \\ &= \partial_a \partial_b w^c + \partial_a (\Gamma^c{}_{fb} w^f) \\ &\quad + \Gamma^c{}_{fa} \partial_b w^f + \Gamma^c{}_{fa} \Gamma^f{}_{gb} w^g - \Gamma^f{}_{ba} \nabla_f w^c \\ &= \partial_a \partial_b w^c + \partial_a \Gamma^c{}_{fb} w^f \\ &\quad + \Gamma^c{}_{fb} \partial_a w^f + \Gamma^c{}_{fa} \partial_b w^f \\ &\quad + \Gamma^c{}_{fa} \Gamma^f{}_{gb} w^g - \Gamma^f{}_{ba} \nabla_f w^c . \end{aligned} \tag{3.36}$$

When we subtract the same expression with a and b reversed, we recognize that for the functions w^c we have $\partial_a \partial_b w^c - \partial_b \partial_a w^c = [e_a, e_b]^d \partial_d w^c$, that the second line $\Gamma^c{}_{fb} \partial_a w^f + \Gamma^c{}_{fa} \partial_b w^f$ vanishes, and that $\Gamma^f{}_{ba} - \Gamma^f{}_{ab} = [e_a, e_b]^f + T^f{}_{ab}$, so that

$$\begin{aligned} \left(\nabla_a \nabla_b - \nabla_b \nabla_a \right) w^c &= [e_a, e_b]^d \partial_d w^c + \partial_a \Gamma^c{}_{fb} w^f - \partial_b \Gamma^c{}_{fa} w^f \\ &\quad + \Gamma^c{}_{fa} \Gamma^f{}_{gb} w^g - \Gamma^c{}_{fb} \Gamma^f{}_{ga} w^g \\ &\quad - \left([e_a, e_b]^f + T^f{}_{ab} \right) \nabla_f w^c , \end{aligned} \tag{3.37}$$

and thus relabeling dummy indices to obtain an expression in terms of w^d, we arrive at

$$\begin{aligned} R^c{}_{dab} w^d &= \left(\nabla_a \nabla_b - \nabla_b \nabla_a + T^d{}_{ab} \nabla_d \right) w^c \\ &= \left(\partial_a \Gamma^c{}_{db} - \partial_b \Gamma^c{}_{da} + \Gamma^c{}_{fa} \Gamma^f{}_{db} - \Gamma^c{}_{fb} \Gamma^f{}_{da} - [e_a, e_b]^f \Gamma^c{}_{df} \right) w^d . \end{aligned} \tag{3.38}$$

This expression follows much more directly from the expression $\check{R} \equiv \mathrm{d}\check{\Gamma} + \check{\Gamma} \wedge \check{\Gamma}$, but the above derivation from the covariant derivative expression is included here to clarify other presentations which are sometimes obscured by the quirks of index notation for covariant derivatives.

> △ The derivation above makes clear how the expression for the curvature in terms of the covariant derivative simplifies to $R^c{}_{dab} w^d = \left(\nabla_a \nabla_b - \nabla_b \nabla_a \right) w^c$ for zero torsion but is unchanged in a holonomic frame, while in contrast the expression in terms of the connection coefficients is unchanged for zero torsion but in a holonomic frame simplifies to omit the term $[e_a, e_b]^f \Gamma^c{}_{df} w^d$.

△ Note that the sign and the order of indices of R as a tensor are not at all consistent across the literature.

3.6 First Bianchi identity

If we take the exterior covariant derivative of the torsion, we get

$$\mathrm{D}\vec{T} = \mathrm{DD}\vec{\beta} = \check{R} \wedge \vec{\beta}. \tag{3.39}$$

This is called the **first** (AKA algebraic) **Bianchi identity**. Using the antisymmetry of $\check{R}$, we can write the first Bianchi identity explicitly as

$$\mathrm{D}\vec{T}(u, v, w) = \check{R}(u, v)\vec{w} + \check{R}(v, w)\vec{u} + \check{R}(w, u)\vec{v}. \tag{3.40}$$

In the case of zero torsion, this identity becomes $\check{R} \wedge \vec{\beta} = 0$, which in index notation can be written ${R^c}_{[dab]} = 0$.

We can find a geometric interpretation for this identity by first constructing a variant of our picture of $\check{R}(u, v)\vec{w}$ as the change in $\vec{w}$ after being parallel transported in opposite directions around a loop. Taking advantage of our previous result that $\check{R}(u, v)\vec{w}$ only depends upon the local values of u and v, we are free to construct their vector field values such that $[u, v] = 0$. We then examine the difference between $\vec{w}$ being parallel transported in each direction halfway around the loop. For infinitesimal parallel transport from a point p along a curve C with tangent v we have $\|_{\varepsilon v} (w|_p) \equiv \|_C (w|_p) = w|_{p+\varepsilon v} - \varepsilon \nabla_v w|_p$. Therefore we find that

$$\begin{aligned} \|_{\varepsilon u}\|_{\varepsilon v} (w|_p) &= \|_u (w|_{p+\varepsilon v} - \varepsilon \nabla_v w|_p) \\ &= w|_{p+\varepsilon v+\varepsilon u} - \varepsilon \nabla_v w\big|_{p+\varepsilon u} - \varepsilon \nabla_u w|_{p+\varepsilon v} + \varepsilon^2 \nabla_u \nabla_v w|_p, \end{aligned} \tag{3.41}$$

so that

$$\begin{aligned} \|_{\varepsilon u}\|_{\varepsilon v} (w|_p) - \|_{\varepsilon v}\|_{\varepsilon u} (w|_p) &= \varepsilon^2 \nabla_u \nabla_v w|_p - \varepsilon^2 \nabla_v \nabla_u w|_p \\ &= \varepsilon^2 \check{R}(u, v)\vec{w}, \end{aligned} \tag{3.42}$$

since $[u, v] = 0$ means that $w|_{p+\varepsilon v+\varepsilon u} = w|_{p+\varepsilon u+\varepsilon v}$. In the case of zero torsion, we can further take advantage of our freedom in choosing the vector field values of u and v by requiring them to equal their parallel transports, i.e. $v|_{p+\varepsilon u} \equiv \|_{\varepsilon u} (v|_p)$ and $u|_{p+\varepsilon v} \equiv \|_{\varepsilon v} (u|_p)$, preserving the property $[u, v] = 0$ due to the vanishing torsion.

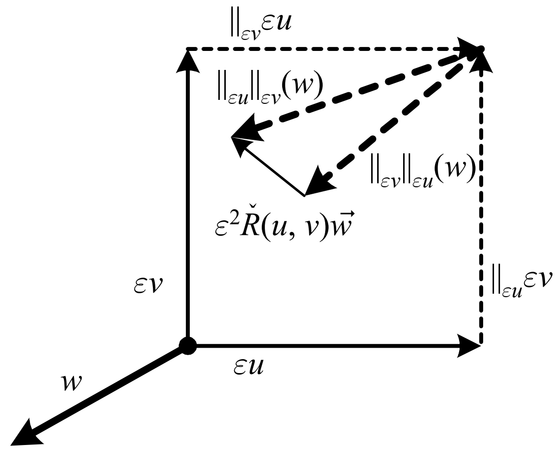

FIGURE 3.7: A slight variant of $\check{R}(u, v)\,\vec{w}$ viewed as "the difference between w when parallel transported around the two opposite edges of the boundary of the surface defined by its arguments." In the case of zero torsion, the boundary can be built from parallel transports instead of vector field values.

Thus, still assuming zero torsion, we can construct a cube from the parallel transports of u, v, and w. This construction reveals that the first Bianchi identity corresponds to the fact that the three curvature vectors form a triangle, i.e. their sum is zero.

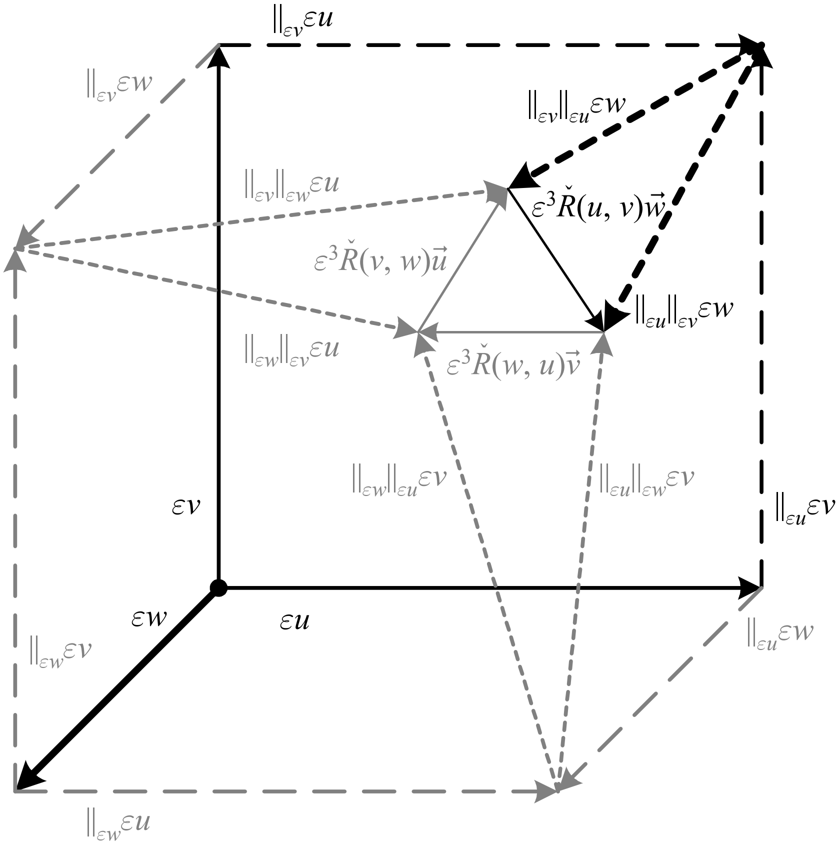

FIGURE 3.8: The first Bianchi identity reflects the fact that for zero torsion, the far corners of a cube made of parallel transported vectors do not meet, and their separation is made up of the differences in parallel transport via opposite edges of each face. Note that the corners of the triangle are points since vanishing torsion means that e.g. $\varepsilon u + \|_{\varepsilon u} (\varepsilon w) = \varepsilon w + \|_{\varepsilon w} (\varepsilon u)$, so that the top point of the triangle reflects this equality parallel transported by εv.

3.7 Second Bianchi identity

If we take the exterior covariant derivative of the curvature, we get

$$\mathrm{D}\check{R} = 0. \tag{3.43}$$

This is called the **second Bianchi identity**, and can be verified algebraically from the definition $\check{R} \equiv \mathrm{d}\check{\Gamma} + \check{\Gamma} \wedge \check{\Gamma}$. We can write this identity more explicitly as

$$\begin{aligned} 0 &= \mathrm{D}\check{R}(u, v, w)\vec{a} \\ &= \nabla_u \check{R}(v, w)\vec{a} + \nabla_v \check{R}(w, u)\vec{a} + \nabla_w \check{R}(u, v)\vec{a} \\ &\quad - \check{R}([u, v], w)\vec{a} - \check{R}([v, w], u)\vec{a} - \check{R}([w, u], v)\vec{a}, \end{aligned} \tag{3.44}$$

where we have used the antisymmetry of $\check{R}$ and the covariant derivative acts on the value of $\check{R}$ as a tensor of type $(1, 1)$. Working this expression into tensor notation and using the tensor expression for the torsion in terms of the commutator, we find that

$$0 = \nabla_e R^c{}_{dab} + \nabla_a R^c{}_{dbe} + \nabla_b R^c{}_{dea} \\ - R^c{}_{dfe} T^f{}_{ab} - R^c{}_{dfa} T^f{}_{be} - R^c{}_{dfb} T^f{}_{ea}, \tag{3.45}$$

or

$$R^c{}_{d[ab;e]} = R^c{}_{df[e} T^f{}_{ab]}, \tag{3.46}$$

and in the case of zero torsion, $R^c{}_{d[ab;e]} = 0$.

Geometrically, the second Bianchi identity can be seen as reflecting the same "boundary of a boundary" idea as that of $\mathrm{d}^2 = 0$ in Fig. C.8, except that here we are parallel transporting a vector $\vec{a}$ around each face that makes up the boundary of the cube. As in the previous section, we can take advantage of the fact that $\check{R}(v, w)\vec{a}$ only depends upon the local value of $\vec{a}$, constructing its vector field values such that e.g. $\vec{a}\,|_{p+\varepsilon u} = \|_{\varepsilon u} (\vec{a}\,|_p)$, giving us

$$\begin{aligned} \varepsilon \nabla_u \check{R}(v, w)\vec{a} &= \check{R}(v\,|_{p+\varepsilon u}, w\,|_{p+\varepsilon u})\vec{a}\,|_{p+\varepsilon u} - \|_{\varepsilon u} \check{R}(v, w) \|_{\varepsilon u}^{-1} \vec{a}\,|_{p+\varepsilon u} \\ &= \check{R}(v\,|_{p+\varepsilon u}, w\,|_{p+\varepsilon u}) \|_{\varepsilon u} \vec{a} - \|_{\varepsilon u} \check{R}(v, w)\vec{a}. \end{aligned} \tag{3.47}$$

The first term parallel transports $\vec{a}$ along εu and then around the parallelogram defined by v and w at $p + \varepsilon u$, while the second parallel transports $\vec{a}$ around the parallelogram defined by v and w at p, then along εu. Thus in the case of vanishing Lie commutators (e.g. a holonomic frame), we construct a cube from the vector fields u, v, and w, and find that the second Bianchi identity reflects the fact that $\mathrm{D}\check{R}(u, v, w)\vec{a}$ parallel transports $\vec{a}$ along each edge of the cube an equal number of times in opposite directions, thus canceling out any changes.

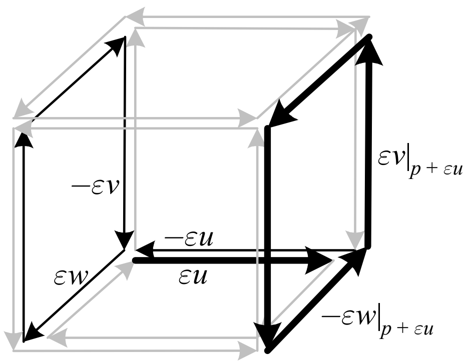

FIGURE 3.9: The second Bianchi identity reflects the fact that for vanishing Lie commutators, $\mathrm{D}\check{R}(u, v, w)\vec{a}$ parallel transports $\vec{a}$ along each edge of the cube made of the three vector field arguments an equal number of times in opposite directions, thus canceling out any changes. Above, $\varepsilon \nabla_u \check{R}(v, w)\vec{a} = \check{R}(v\,|_{p+\varepsilon u}, w\,|_{p+\varepsilon u}) \|_{\varepsilon u} \vec{a} - \|_{\varepsilon u} \check{R}(v, w)\vec{a}$ is highlighted by the bold arrows representing the path along which $\vec{a}$ is parallel transported in the first term, and by the remaining dark arrows representing the path along which $\vec{a}$ is parallel transported in the second term.

In the case of non-vanishing torsion, where there is a non-vanishing commutator $\vec{T}(u, v) = -[u, v] \neq 0$, we find that the cube gains a "shaved edge," and that the extra non-vanishing term $-\check{R}([u, v], w)\vec{a}$ in $\mathrm{D}\check{R}$ maintains the "boundary of a boundary" logic by adding a loop of parallel transports of $\vec{a}$ in the proper direction around the new "face" created.

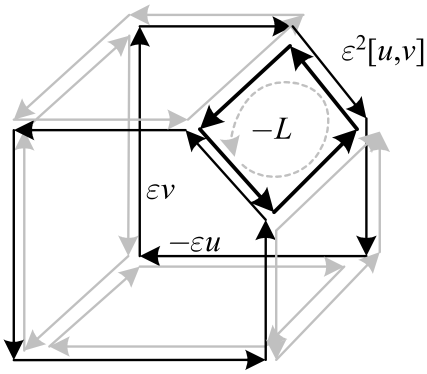

FIGURE 3.10: In the case of non-vanishing torsion and thus commutator, the extra term $-\check{R}([u,v],w)\vec{a}$ in $\mathrm{D}\check{R}$ maintains the cancellation of face boundaries by adding a loop L around the new "shaved edge" created.

4 Introducing the metric

4.1 The Riemannian metric

A (pseudo) metric tensor (see Section A.4) is a (pseudo) inner product $\langle v, w\rangle$ on a vector space V that can be represented by a symmetric tensor g_{ab}, and thus can be used to lower and raise indices on tensors. A **(pseudo) Riemannian metric** (AKA metric) is a (pseudo) metric tensor field on a manifold M, making M a **(pseudo) Riemannian manifold**.

A metric defines the length (norm) of tangent vectors, and can thus be used to define the length L of a curve C via parametrization and integration:

$$\begin{aligned} L(C) &\equiv \int \left\| \dot{C}(t) \right\| \mathrm{d}t \\ &= \int \sqrt{\left\langle \dot{C}(t), \dot{C}(t) \right\rangle} \mathrm{d}t \end{aligned} \tag{4.1}$$

This also turns any (non-pseudo) Riemannian manifold into a metric space, with distance function $d(x, y)$ defined to be the minimum length curve connecting the two points x and y; this curve is always a geodesic, and any geodesic locally minimizes the distance between its points (only locally since e.g. a geodesic may eventually self-intersect as the equator on a sphere does).

☼ With a metric, our intuitive picture of a manifold loses its "stretchiness" via the introduction of length and angles; but having only intrinsically defined properties, the manifold can still be e.g. rolled up like a piece of paper if imagined as flat and embedded in a larger space.

If the coordinate frame of x^μ is orthonormal at a point $p \in M^n$ in a Riemannian manifold, for arbitrary coordinates y^μ we can consider the components of the metric tensor in the two coordinate frames to find that

$$\begin{aligned} g_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\nu &= \delta_{\lambda\sigma}\mathrm{d}x^\lambda \mathrm{d}x^\sigma \\ &= \delta_{\lambda\sigma}\frac{\partial x^\lambda}{\partial y^\mu}\mathrm{d}y^\mu \frac{\partial x^\sigma}{\partial y^\nu}\mathrm{d}y^\nu \\ &= [J_x(y)]^T [J_x(y)]\,\mathrm{d}y^\mu \mathrm{d}y^\nu \\ \Rightarrow \det(g_{\mu\nu}) &= [\det(J_x(y))]^2, \end{aligned} \tag{4.2}$$

where $J_x(y)$ is the Jacobian matrix (see Section B.6) and we have used the fact that $\det(A^T A) = [\det(A)]^2$. Thus the volume of an region $U \in M^n$ corresponding to $R \in \mathbb{R}^n$ in the coordinates x^μ is

$$V(U) = \int_R \sqrt{\det(g)}\mathrm{d}x^1 \ldots \mathrm{d}x^n, \tag{4.3}$$

where $\det(g)$ is the determinant of the metric tensor as a matrix in the coordinate frame $\partial/\partial x^\mu$. In the context of a pseudo-Riemannian manifold $\det(g)$ can be negative, and the integrand

$$\mathrm{d}V \equiv \sqrt{|\det(g)|}\mathrm{d}x^1 \ldots \mathrm{d}x^n \tag{4.4}$$

is called the **volume element**, or when written as a form $\mathrm{d}V \equiv \sqrt{|\det(g)|}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ it is called the **volume form**. In physical applications $\mathrm{d}V$ usually denotes the **volume pseudo-form**, which gives a positive value regardless of orientation. Note that if the coordinate basis is orthonormal then $|\det(g)| = 1$; thus these definitions are consistent with those typically defined on $\mathbb{R}^n$. Sometimes one defines a volume form on a manifold without defining a metric; in this case the metric (and connection) is not uniquely determined.

> △ The symbol g is frequently used to denote $\det(g)$, and sometimes $\sqrt{|\det(g)|}$, in addition to denoting the metric tensor itself.

We can use the inner product to define an **orthonormal frame** on M. In four dimensions an orthonormal frame is also called a **tetrad** (AKA vierbein). Any frame on a manifold can be defined to be an orthonormal frame, which is equivalent to defining the metric (which in the orthonormal frame is $g_{ab} = \eta_{ab}$). An orthonormal holonomic frame exists on a region of M if and only if that region is flat. Thus in general, given a set of coordinates on M, we have to choose between using either a non-coordinate orthonormal frame or a non-orthonormal coordinate frame.

The **Hopf-Rinow theorem** says that a connected Riemannian manifold M is complete as a metric space (or equivalently, all closed and bounded subsets are compact) if and only if it is **geodesically complete**, meaning that the exponential map is defined for all vectors at some $p \in M$. If M is geodesically complete at p, then it is at all points on the manifold, so this property can also be used to state the theorem. This theorem is not valid for pseudo-Riemannian manifolds; any (pseudo) Riemannian manifold that is geodesically complete is called a **geodesic manifold**.

As noted previously, a Riemannian metric can be defined on any differentiable manifold. In general, not every manifold admits a pseudo-Riemannian metric, and in particular not every 4-manifold admits a Minkowski metric, but 4-manifolds that are noncompact, parallelizable, or compact, connected and of Euler characteristic 0 all do.

In the same way that differentiable manifolds are equivalent if they are related by a diffeomorphism, Riemannian manifolds are equivalent if they are related by an **isometry**, a diffeomorphism $\Phi\colon M \to N$ that preserves the metric, i.e. $\forall v, w \in TM$, $\langle v, w\rangle|_p = \langle \mathrm{d}\Phi_p(v), \mathrm{d}\Phi_p(w)\rangle\big|_{\Phi(p)}$. Also like diffeomorphisms, the isometries of a manifold form a group; for example, the group of isometries of Minkowski space is the Poincarï¿½ group. A vector field whose one-parameter diffeomorphisms are isometries is called a **Killing field**, also called a **Killing vector** since it can be shown ([8] pp. 188-189) that a Killing field is determined by a vector at a single point along with its covariant derivatives. A Killing field thus satisfies $L_v g_{ab} = 0$, which using eq. (3.26) for a Levi-Civita connection (see next section) is equivalent to

$$\nabla_a v_b + \nabla_b v_a = 0, \tag{4.5}$$

called the **Killing equation** (AKA Killing condition).

We can then consider isometric immersions and embeddings, and ask whether every Riemannian manifold can be embedded in some $\mathbb{R}^n$. The **Nash embedding theorem** provides an affirmative answer, and it can also be shown that every pseudo-Riemannian manifold can be isometrically embedded in some $\mathbb{R}^n$ with some signature while maintaining arbitrary differentiability of the metric.

4.2 The Levi-Civita connection

A connection on a Riemannian manifold M is called a **metric connection** (AKA metric compatible connection, isometric connection) if its associated parallel transport respects the metric, i.e. it preserves lengths and angles. More precisely, $\forall v, w \in TM$, we require that

$$\langle \|_C\,(v), \|_C\,(w)\rangle = \langle v, w\rangle \tag{4.6}$$

for any curve C in M.

In terms of the metric, this can be written $g_{ab} \|_C\, v^a \|_C\, w^b = g_{ab} v^a w^b$. But recalling that the parallel transport of tensors just transports the arguments, we also have $(\|_{-C}\, g_{ab})\, v^a w^b = g_{ab} \|_C\, v^a \|_C\, w^b$, so that we must have $\|_{-C}\, g_{ab} = g_{ab}$, or $\nabla_c g_{ab} = 0$. In terms of the connection coefficients, a metric connection then satisfies

$$\nabla_c g_{ab} = \partial_c g_{ab} - \Gamma^d{}_{ac} g_{db} - \Gamma^d{}_{bc} g_{ad} = 0. \tag{4.7}$$

Using the Leibniz rule for the covariant derivative over the tensor product, we can derive a Leibniz rule over the inner product:

$$\begin{aligned} \nabla_c \left(g_{ab} v^a w^b \right) &= 0 + g_{ab} \nabla_c v^a w^b + g_{ab} v^a \nabla_c w^b \\ \Rightarrow \nabla_u \langle v, w\rangle &= \langle \nabla_u v, w\rangle + \langle v, \nabla_u w\rangle \end{aligned} \tag{4.8}$$

Requiring this relationship to hold is an equivalent way to define a metric connection.

The **Levi-Civita connection** (AKA Riemannian connection, Christoffel connection) is then the torsion-free metric connection on a (pseudo) Riemannian manifold M. The **fundamental theorem of Riemannian geometry** states that for any (pseudo) Riemannian manifold the Levi-Civita connection exists and is unique. On the other hand, an arbitrary connection can only be the Levi-Civita connection for some metric if it is torsion-free and preserves lengths; moreover, this metric is unique only up to a scaling factor (in physics, this corresponds to a choice of units).

For a metric connection, the curvature then must take values that are infinitesimal rotations, i.e. $\check{R}$ is $o(r,s)$-valued. Thus if we eliminate the influence of the signature by lowering the first index, the first two indices of the curvature tensor are anti-symmetric:

$$R_{cdab} = -R_{dcab} \tag{4.9}$$

Using the anti-symmetry of the other indices and the first Bianchi identity, this leads to another commonly noted symmetry

$$R_{cdab} = R_{abcd}. \tag{4.10}$$

The Leibniz rule for the covariant derivative over the inner product along with the zero torsion relation $\nabla_v w = \nabla_w v + [v, w]$ can be used to derive an expression called the **Koszul formula**:

$$\begin{aligned} 2\langle \nabla_u v, w\rangle =& \nabla_u \langle v, w\rangle + \nabla_v \langle w, u\rangle - \nabla_w \langle u, v\rangle \\ &- \langle u, [v, w]\rangle + \langle v, [w, u]\rangle + \langle w, [u, v]\rangle \end{aligned} \tag{4.11}$$

Substituting in the frame vector fields and eliminating the metric tensor from the left hand side, we arrive at an expression for the connection in terms of the metric:

$$\begin{aligned} 2\Gamma^{c}{}_{ba} = g^{cd}(&\partial_a g_{bd} + \partial_b g_{da} - \partial_d g_{ab} \\ &- g_{af}[e_b, e_d]^f + g_{bf}[e_d, e_a]^f + g_{df}[e_a, e_b]^f) \end{aligned} \tag{4.12}$$

On a Riemannian manifold, the connection coefficients for the Levi-Civita connection in a coordinate basis $\Gamma^{\lambda}{}_{\mu\sigma}$ are called the **Christoffel symbols**. At a point $p \in U \subset M$, an orthonormal basis for T_pU can be used to form geodesic normal coordinates, which are then called **Riemann normal coordinates**. Recalling from Section 2.6 that with zero torsion the connection coefficients vanish at p, we can apply the covariant derivative to the metric tensor to conclude that the partial derivatives of the metric $g_{\mu\nu} = \eta_{\mu\nu}$ all also vanish at p.

☼ The vanishing of the Christoffel symbols at the origin of Riemann normal coordinates is frequently used to simplify the derivation of tensor relations which are then, being frame-independent, seen to be true in any coordinate system or frame (and if the origin was chosen arbitrarily, at any point). In particular, the covariant and partial derivatives are equivalent at the origin of Riemann normal coordinates.

4.3 Independent quantities and dependencies

While in general the curvature on a Riemannian manifold does not determine the metric, for a manifold with connection that is compact, simply connected, and has no regions of constant curvature (i.e. there is no way to "stretch" the manifold without affecting the curvature), knowledge of the curvature at all points determines the connection (up to changes in frame), and therefore the metric that makes this connection Levi-Civita (up to a constant scaling factor).

If we choose coordinates and use a coordinate frame on M^n, we can calculate the number of independent functions and equations associated with the various quantities and relations we have covered, and use them to verify the associated dependencies.

Quantity / relation	Viewpoint	Count
Metric	Symmetric matrix of functions	$n(n+1)/2$
Coordinate frame	Fixed	0
Connection	gl-valued (matrix-valued) 1-form	n^3
Metric condition	Derivative of metric	$n^2(n+1)/2$
Torsion-free condition	Vector-valued 2-form	$n^2(n-1)/2$

Table 4.1: Independent function and equation counts in a coordinate frame.

The choice of coordinates determines the frame, leaving the geometry of the Riemannian manifold defined by the $n(n+1)/2$ functions of the metric. A torsion-free connection consists of $n^3 - n^2(n-1)/2 = n^2(n+1)/2$ functions. The metric condition is exactly this number of equations, allowing us in general to solve for the connection if the metric is known, or vice-versa (up to a constant scaling factor).

Alternatively, we can look at things in a orthonormal frame:

Quantity / relation	Viewpoint	Count
Metric	Fixed	0
Orthonormal frame	n vector fields	n^2
Change of orthonormal frame	SO-valued 0-form	$n(n-1)/2$
Connection	so-valued 1-form	$n^2(n-1)/2$
Metric condition	Automatically satisfied	0
Torsion-free condition	Vector-valued 2-form	$n^2(n-1)/2$

Table 4.2: Independent function and equation counts in an orthonormal frame.

Here the metric is constant, and the orthonormal frame consists of n^2 functions, but it is determined only up to a change of orthonormal frame (rotation), leaving $n^2 - n(n-1)/2 = n(n+1)/2$ functions, consistent with the metric function count above. The torsion-free condition is the same number of equations as the connection has functions, so that in general the torsion-free connection can be determined by the orthonormal frame.

4.4 The divergence and conserved quantities

The divergence of a vector field u (see Section C.5) can be generalized to a pseudo-Riemannian manifold (sometimes called the **covariant divergence**) by defining

$$\mathrm{div}(u) \equiv (-1)^s * \mathrm{d}(*(u^\flat)). \tag{4.13}$$

Using the relations $i_u\Omega = (-1)^s * (u^\flat)$ (see Section C.6) and $A = (*A)\Omega$ for $A \in \Lambda^n M^n$ (see Section A.10), we have

$$\begin{aligned} \mathrm{d}(i_u\Omega) &= (-1)^s \mathrm{d}(*(u^\flat)) \\ &= (-1)^s * \mathrm{d}(*(u^\flat))\Omega \\ &= \mathrm{div}(u)\Omega. \end{aligned} \tag{4.14}$$

Using $i_u\mathrm{d} + \mathrm{d}i_u = L_u$ we then arrive at $\mathrm{div}(u)\Omega = L_u\Omega$, or as it is more commonly written

$$\mathrm{div}(u)\mathrm{d}V = L_u\mathrm{d}V. \tag{4.15}$$

Thus we can say that $\text{div}(u)$ is "the fraction by which a unit volume changes when transported by the flow of u," and if $\text{div}(u) = 0$ then we can say that "the flow of u leaves volumes unchanged." Expanding the volume element in coordinates x^λ we can obtain an expression for the divergence in terms of these coordinates,

$$\text{div}(u) = \frac{1}{\sqrt{|\det(g)|}} \partial_\lambda \left(u^\lambda \sqrt{|\det(g)|} \right). \tag{4.16}$$

Note that both this expression and $\nabla_a u^a$ are coordinate-independent and equal to $\partial_a u^a$ in Riemann normal coordinates, confirming our expectation that in general we have

$$\text{div}(u) = \nabla_a u^a. \tag{4.17}$$

Using the relation $\text{div}(u)\Omega = \text{d}(i_u \Omega)$ from eq. (4.14), along with Stokes' theorem, we recover the classical **divergence theorem**

$$\begin{aligned} \int_V \text{div}(u) \text{d}V &= \int_{\partial V} i_u \text{d}V \\ &= \int_{\partial V} \langle u, n \rangle \, \text{d}S, \end{aligned} \tag{4.18}$$

where V is an n-dimensional compact submanifold of M^n, n is the unit normal vector to ∂V, and $\text{d}S \equiv i_n \text{d}V$ is the induced volume element ("surface element") for ∂V. In the case of a Riemannian metric, this can be thought of as reflecting the intuitive fact that "the change in a volume due to the flow of u is equal to the net flow across that volume's boundary." If $\text{div}(u) = 0$ then we can say that "the net flow of u across the boundary of a volume is zero." We can also consider an infinitesimal V, so that the divergence at a point measures "the net flow of u across the boundary of an infinitesimal volume." As usual, for a pseudo-Riemannian metric these geometric intuitions have less meaning.

The divergence can be extended to contravariant tensors T by defining

$$\text{div}(T) \equiv \nabla_a T^{ab}, \tag{4.19}$$

although other conventions are in use. Since $\text{div}(T)$ is vector-valued and the parallel transport of vectors is path-dependent, we cannot in general integrate to get a divergence theorem for tensors. In the case of a flat metric however, we are able to integrate to get a divergence theorem for each component

$$\int_V \nabla_a T^{ab} \text{d}V = \int_{\partial V} T_a{}^b n^a \text{d}S. \tag{4.20}$$

In physics, the vector field u often represents the **current vector** (AKA current density, flux, flux density) $j \equiv \rho u$ of an actual physical flow, where ρ is the density of the physical quantity Q and u is thus a velocity field; e.g. in $\mathbb{R}^3$, j has units $Q/(\text{length})^2(\text{time})$. There are several quantities that can be defined around this concept:

Quantity	Definition	Meaning
Current vector	$j \equiv \rho u$	The vector whose length is the amount of Q per unit time crossing a unit area perpendicular to j
Current form	$\zeta \equiv i_j \mathrm{d}V$ $= \langle j, n \rangle \, \mathrm{d}S$	The $(n-1)$-form which gives the amount of Q per unit time crossing the area defined by the argument vectors
Current density	$\mathrm{j} \equiv \sqrt{\lvert\det(g)\rvert}\, j$ $\Rightarrow \zeta = \langle \mathrm{j}, n \rangle \, \mathrm{d}x^{\lambda_1} \wedge \cdots \wedge \mathrm{d}x^{\lambda_{n-1}}$	The vector whose length is the amount of Q per unit time crossing a unit coordinate area perpendicular to j
Current	$I \equiv \int_S \zeta$ $= \int_S \langle j, n \rangle \, \mathrm{d}S$ $= \int_{S(x^\lambda)} \langle \mathrm{j}, n \rangle \, \mathrm{d}x^{\lambda_1} \cdots \mathrm{d}x^{\lambda_{n-1}}$	The amount of Q per unit time crossing S
Current 4-vector	$J \equiv (\rho, j^\mu)$	Current vector on the spacetime manifold

TABLE 4.3: Quantities related to current. ρ is the density of the physical quantity Q, u is a velocity field, n is the unit normal to a surface S, and x^λ are coordinates on the submanifold S. The current 4-vector can be generalized to other Lorentzian manifolds, and can also be turned into a form or a density.

△ Note that the terms flux and current (as well as flux density and current density) are not used consistently in the literature.

The current density j is an example of a **tensor density**, which in general takes the form

$$\mathfrak{T} \equiv \left(\sqrt{\lvert\det(g)\rvert}\right)^W T, \tag{4.21}$$

where T is a tensor and W is called the **weight**. Note that tensor densities are not coordinate-independent quantities.

For a Riemannian metric we now define the **continuity equation** (AKA equation of continuity)

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \Sigma - \int_{\partial V} \langle j, n \rangle \, \mathrm{d}S, \tag{4.22}$$

where q is the amount of Q contained in V, t is time, and Σ is the rate of Q being created within V. The continuity equation thus states the intuitive fact that the change of Q within V equals the amount generated less the amount which passes through ∂V. Using the divergence theorem, we can then obtain the differential form of the continuity equation

$$\frac{\partial \rho}{\partial t} = \sigma - \mathrm{div}(j), \tag{4.23}$$

where σ is the amount of Q generated per unit volume per unit time. This equation then states the intuitive fact that at a point, the change in density of Q equals the amount generated less the amount that moves away. Positive σ is referred to as a **source** of Q, and negative σ a **sink**. If $\sigma = 0$ then we say that Q is a **conserved quantity** and refer to the continuity equation as a (local) **conservation law**.

Under a flat Lorentzian metric, we can combine ρ and j into the current 4-vector J and express the continuity equation with $\sigma = 0$ as

$$\mathrm{div}(J) = 0, \tag{4.24}$$

whereupon J is called a **conserved current**. Note that in this approach we lose the intuitive meaning of the divergence under a Riemannian metric. If any curvature is present, when we split out the time component we recover a Riemannian divergence but introduce a source due to the non-zero Christoffel symbols

$$\begin{aligned} \nabla_\mu J^\mu &= \partial_\mu J^\mu + \Gamma^\mu{}_{\nu\mu} J^\nu \\ &= \partial_t \rho + \nabla_i j^i + \left(\Gamma^\mu{}_{t\mu} \rho + \Gamma^t{}_{it} j^i \right), \end{aligned} \tag{4.25}$$

where t is the negative signature component and the index i goes over the remaining positive signature components. Thus, since the Christoffel symbols are coordinate-dependent, in the presence of curvature there is in general no coordinate-independent conserved quantity associated with a vanishing Lorentzian divergence.

4.5 Ricci and sectional curvature

The **Ricci curvature tensor** (AKA Ricci tensor) is formed by contracting two indices in the Riemann curvature tensor:

$$\begin{aligned} R_{ab} &\equiv R^c{}_{acb} \\ \mathrm{Ric}(v, w) &\equiv R_{ab} v^a w^b \end{aligned} \tag{4.26}$$

Using the symmetries of the Riemann tensor for a metric connection along with the first Bianchi identity for zero torsion, it is easily shown that the Ricci tensor is symmetric.

Since the Ricci tensor is symmetric, by the spectral theorem it can be diagonalized and thus is determined by

$$\mathrm{Ric}(v) \equiv \mathrm{Ric}(v, v), \tag{4.27}$$

which is called the **Ricci curvature function** (AKA Ricci function). Note that the Ricci function is not a 1-form since it is not linear in v. Choosing a basis that diagonalizes R_{ab} is equivalent to choosing our basis vectors to line up with the directions that yield extremal values of the Ricci function on the unit vectors $\mathrm{Ric}(\hat{v}, \hat{v})$ (or equivalently, the principal axes of the ellipsoid / hyperboloid $\mathrm{Ric}(v, v) = 1$).

Finally, if we raise one of the indices of the Ricci tensor and contract we arrive at the **Ricci scalar** (AKA scalar curvature):

$$R \equiv g^{ab} R_{ab} \tag{4.28}$$

For a Riemannian manifold M^n, the Ricci scalar can thus be viewed as n times the average of the Ricci function on the set of unit tangent vectors.

△ The Ricci function and Ricci scalar are sometimes defined as averages instead of contractions (sums), introducing extra factors in terms of the dimension n to the above definitions.

The Ricci function in terms of the curvature 2-form in an orthonormal frame e_μ (dropping the hats to avoid clutter) on a pseudo-Riemannian manifold M^n naturally splits into terms which each also measure curvature:

$$\mathrm{Ric}(e_\mu) = \sum_{i \neq \mu} g_{ii} \left\langle \check{R}(e_i, e_\mu)\vec{e}_\mu, e_i \right\rangle \tag{4.29}$$

The term $i = \mu$ vanishes due to the anti-symmetry of $\check{R}$. The $(n-1)$ non-zero terms are each called a **sectional curvature**, which in general is defined as

$$\begin{aligned} K(v, w) &\equiv \frac{\left\langle \check{R}(v, w)\vec{w}, v \right\rangle}{\langle v, v\rangle \langle w, w\rangle - \langle v, w\rangle^2} \\ \Rightarrow K(e_i, e_j) &= g_{ii} g_{jj} \left\langle \check{R}(e_i, e_j)\vec{e}_j, e_i \right\rangle \\ \Rightarrow \mathrm{Ric}(e_\mu) &= \sum_{i \neq \mu} g_{\mu\mu} K(e_i, e_\mu) \\ \Rightarrow R &= \sum_j g_{jj} \mathrm{Ric}(e_j) \\ &= \sum_{i \neq j} K(e_i, e_j) \\ &= 2 \sum_{i < j} K(e_i, e_j). \end{aligned} \tag{4.30}$$

Note that the sectional curvature is not a 2-form since it is not linear in its arguments; in fact it is constructed to only depend on the plane defined by them, and therefore is symmetric and defined to vanish for equal arguments. Thus for a Riemannian manifold, the Ricci function of a unit vector $\mathrm{Ric}(\hat{v})$ can be viewed as $(n-1)$ times the average of the sectional curvatures of the planes that include $\hat{v}$, and the Ricci scalar can be viewed as n times the average of all the Ricci functions. For a pseudo-Riemannian manifold, the Ricci scalar is twice the sum of all sectional curvatures, or $n(n-1)$ times the average of all sectional curvatures, whose count is the binomial coefficient n choose 2 or $n(n-1)/2$.

The sectional curvatures completely determine the Riemann tensor, but in general the Ricci tensor alone does not for manifolds of dimension greater than 3. However, the Riemann tensor is determined by the Ricci tensor together with the **Weyl curvature tensor** (AKA Weyl tensor, conformal tensor), whose definition (not reproduced here) removes all contractions of the Riemann tensor, so that it is the "trace-free part of the curvature" (i.e. all of its contractions vanish). The Weyl tensor is only defined and non-zero for dimensions $n > 3$.

The **Einstein tensor** is defined as

$$\begin{aligned} G(v, w) &\equiv \mathrm{Ric}(v, w) - \frac{R}{2} g(v, w) \\ G_{ab} &= R_{ab} - \frac{R}{2} g_{ab}. \end{aligned} \tag{4.31}$$

If we define $G \equiv g^{ab}G_{ab}$ then we find that $R_{ab} = G_{ab} - Gg_{ab}/(n-2)$, so that the Einstein tensor vanishes iff the Ricci tensor does. Now, the Einstein tensor is symmetric, and by the spectral theorem can be diagonalized at a given point in an orthonormal basis, which also diagonalizes the Ricci tensor. In terms of the sectional curvature, we have

$$G(e_\mu, e_\mu) = -g_{\mu\mu} \sum_{\substack{i<j \\ i,j\neq\mu}} K(e_i, e_j). \tag{4.32}$$

Thus for a Riemannian manifold, the Einstein tensor $G(\hat{v}, \hat{v})$ applied to a unit vector twice can be viewed as $-\langle \hat{v}, \hat{v}\rangle (n-1)(n-2)/2$ times the average of the sectional curvatures of the planes orthogonal to $\hat{v}$. Using the second Bianchi identity it can be shown ([3] pp. 80-81) that the Einstein tensor is also "divergenceless," i.e.

$$\nabla_a G^{ab} = 0. \tag{4.33}$$

Recall that unless the metric is flat, there is no conserved quantity which can be associated with this vanishing divergence.

△ Frequent references to the divergencelessness of the Einstein tensor being related to a conserved quantity usually refer to some kind of particular context; one simple one is that in the limit of zero curvature, there is a set of conserved quantities due to eq. (4.20).

4.6 Curvature and geodesics

Geometrically, the Ricci function $\mathrm{Ric}(v)$ at a point $p \in M^n$ can be seen to measure the extent to which the area defined by the geodesics emanating from the $(n-1)$-surface perpendicular to v changes in the direction of v. Considering the three dimensional case in an orthonormal frame(and again dropping the hats in $\hat{e}_i$ to avoid clutter), we have

$$\begin{aligned} \mathrm{Ric}(e_2) &= \left\langle \check{R}(e_1, e_2)\vec{e}_2, e_1 \right\rangle + \left\langle \check{R}(e_3, e_2)\vec{e}_2, e_3 \right\rangle \\ &= K(e_1, e_2) + K(e_3, e_2). \end{aligned} \tag{4.34}$$

If we form a cube made from parallel transported vectors as we did for the first Bianchi identity, we can see that each sectional curvature term in $\mathrm{Ric}(e_2)$ takes an edge of the cube and measures the length of the difference between the cube-aligned component of its parallel transport in the e_2 direction and the edge of the cube at a point parallel transported in the e_2 direction.

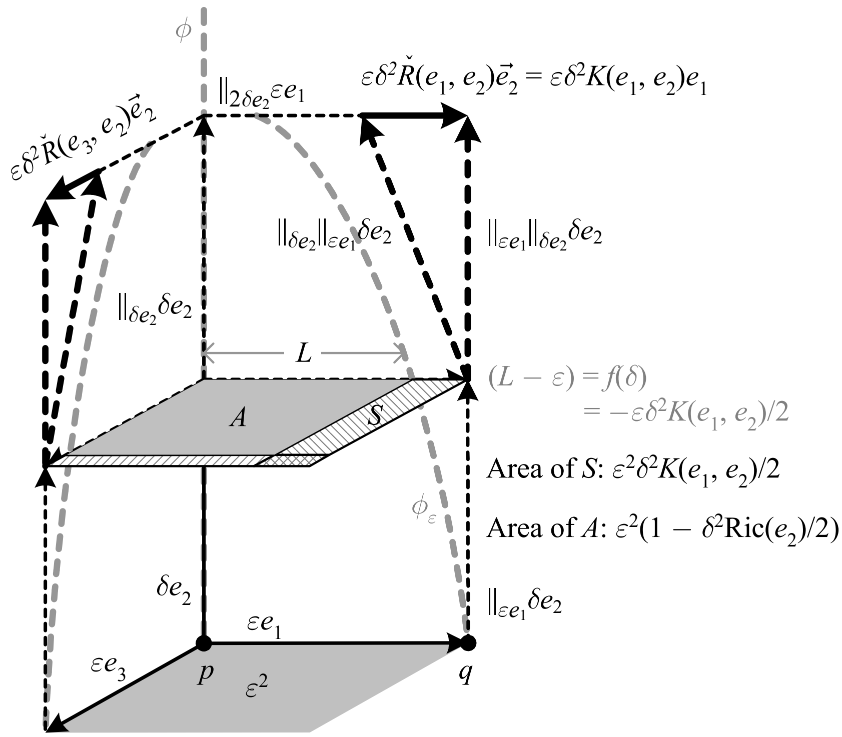

FIGURE 4.1: Each sectional curvature measures the convergence of geodesics, while their sum forms the Ricci curvature function, which measures the change in the area of the $(n-1)$-surface formed by geodesics perpendicular to its argument. In the figure we assume without loss of generality (see below) that $\check{R}(e_1, e_2)\vec{e}_2$ is parallel to e_1.

The figure above details the sectional curvature $K(e_1, e_2) = \beta^1 \check{R}(e_1, e_2)\vec{e}_2$ assuming that $\check{R}(e_1, e_2)\vec{e}_2$ is parallel to e_1, so that $\langle \check{R}(e_1, e_2)\vec{e}_2, e_1\rangle = \|\check{R}(e_1, e_2)\vec{e}_2\|$. The parallel transport of e_2 along itself is depicted as parallel, so that the geodesic parametrized by arclength $\phi(t)$ is a straight line in the figure. The vector $\|_{\delta e_2}\|_{\varepsilon e_1}\ \delta e_2$ is the parallel transport of $\|_{\varepsilon e_1}\ \delta e_2$ by δ in the direction parallel to e_2, and therefore the geodesic $\phi_\varepsilon(t)$ tangent to $\|_{\varepsilon e_1}\ \delta e_2$ at q has tangent $\|_{\delta e_2}\|_{\varepsilon e_1}\ \delta e_2$ after moving a distance δ. If we consider the function $f(t)$ whose value at $t = \delta$ is the quantity $(L - \varepsilon)$ in the figure (i.e. $f(t)$ measures the offset of the geodesic from the right edge of the stack of parallel cubes), its derivative is the slope of the tangent, so that to lowest order in t we have

$$\begin{aligned} \dot{f}(t) &= -\varepsilon t^2 K(e_1, e_2)/t \\ &= -\varepsilon t K(e_1, e_2) \\ \Rightarrow f(t) &= -\varepsilon t^2 K(e_1, e_2)/2. \end{aligned} \tag{4.35}$$

We can generalize this logic to arbitrary unit vectors $\hat{v}$ and $\hat{w}$ to conclude that $K(\hat{v}, \hat{w})/2$ is the "fraction by which the geodesic parallel to $\hat{w}$ starting $\hat{v}$ away bends towards $\hat{w}$." More precisely, in terms of the distance function and the exponential map, to order ε and δ^2 we have

$$d\left(\exp(\delta\hat{w}), \exp(\delta\ \|_{\varepsilon\hat{v}}\ \hat{w})\right) = \varepsilon\left(1 - \frac{\delta^2}{2}K(\hat{v}, \hat{w})\right). \tag{4.36}$$

In the general case L in the figure is the distance between two geodesics infinitesimally separated in the $\hat{v}$ direction, so if we define $L(t)$ as this distance at any point along the parametrized geodesic tangent to $\hat{w}$, the above becomes

$$L(t) = L(0)\left(1 - \frac{t^2}{2}K(\hat{v},\hat{w})\right)$$
$$\Rightarrow \frac{\ddot{L}(t)}{L(t)} = -K(\hat{v},\hat{w}),$$

where the double dots indicate the second derivative with respect to t. Thus $K(\hat{v},\hat{w})$ is "the acceleration of two parallel geodesics in the $\hat{w}$ direction with initial separation $\hat{v}$ towards each other as a fraction of the initial gap."

Now, the distance $|\varepsilon - L| = \varepsilon\delta^2 K(e_1, e_2)/2$ defines a strip S bordering the surface orthogonal to e_2 a distance δ in the e_2 direction. This strip thus has an area $\varepsilon^2\delta^2 K(e_1, e_2)/2$. If we sum this with the other strip of area $\varepsilon^2\delta^2 K(e_3, e_2)/2$, to order ε^2 and δ^2 we measure the extent to which the area A defined by the geodesics emanating from the surface perpendicular to e_2 changes in the direction of e_2. But the sum of sectional curvatures is just the Ricci function, so that in general $\mathrm{Ric}(v)/2$ is the "fraction by which the area defined by the parallel geodesics emanating from the $(n-1)$-surface perpendicular to v changes in the direction of v." More precisely, we can follow the same logic as above, defining the "infinitesimal geodesic $(n-1)$-area" $A(t)$ along a parametrized geodesic tangent to v, so that to order ε^2 and t^2 we have

$$A(t) = \varepsilon^2\left(1 - \frac{t^2}{2}\mathrm{Ric}(v)\right)$$
$$\Rightarrow \frac{\ddot{A}(t)}{A(t)} = -\mathrm{Ric}(v).$$

Thus $\mathrm{Ric}(v)$ is "the acceleration of the parallel geodesics emanating from the $(n-1)$-surface perpendicular to v towards each other as a fraction of the initial surface." Note that if our previous assumption that $\check{R}(e_1, e_2)\vec{e}_2$ is parallel to e_1 is dropped, the only impact is that of an e_3 component on the area calculation; to address this, a more accurate picture would be to extend the area to include all four quadrants defined by both negative and positive values of e_1 and e_3, in which case any change in area due to an e_3 component cancels. In the case of a pseudo-Riemannian manifold, "areas" and "volumes" become less geometric concepts; however, we have a clear picture in the case of a Lorentzian manifold that the Ricci function applied to a time-like vector $v \equiv \partial/\partial x^0 = \partial/\partial t$ tells us how the infinitesimal space-like volume V of free-falling particles (i.e. following geodesics) changes over time according to

$$\begin{aligned}\frac{\ddot{V}(t)}{V(t)} &= -\mathrm{Ric}(v) \\ &= -R_{00} \\ &= -R^{\mu}{}_{0\mu0}.\end{aligned} \tag{4.37}$$

4.7 Jacobi fields and volumes

Now let us consider a vector field $J(t)$ along the geodesic $\phi(t)$ such that $J(0) \equiv J\,|_p = J\,\big|_{\phi(0)} = e_1$ and $J(\delta) \equiv J\,\big|_{\phi(\delta)} = (L/\varepsilon)\,\|_{\delta e_2}\, e_1$, i.e. J is the vector field "between adjacent geodesics."

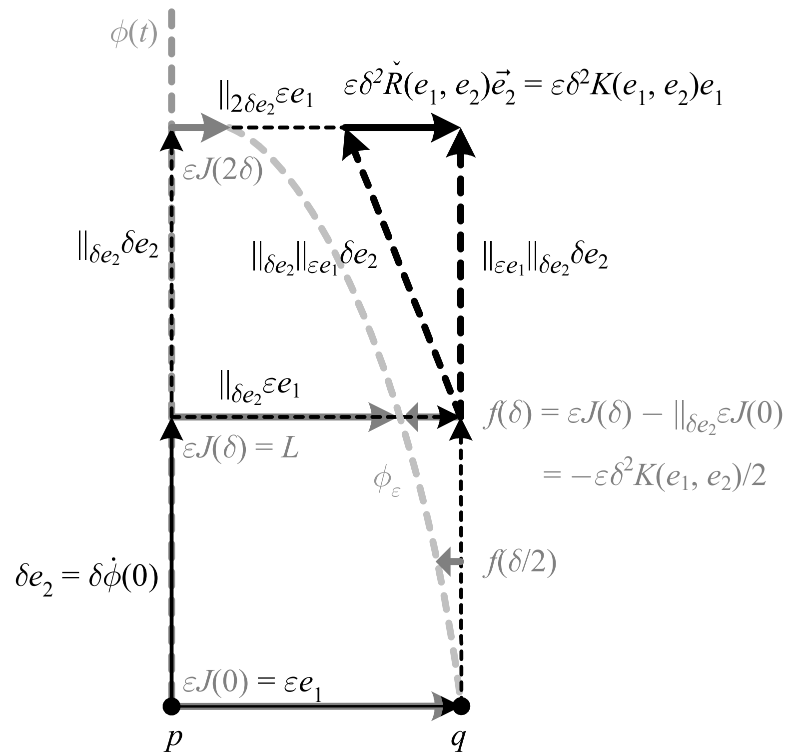

FIGURE 4.2: A Jacobi field is the vector field between adjacent geodesics, whose construction creates a relationship between the covariant derivative and the sectional curvature.

Then the function

$$\begin{aligned} f(t) &= -\varepsilon t^2 K(e_1, e_2)/2 \\ &= -\varepsilon t^2 K(J, \dot{\phi})/2 \end{aligned} \tag{4.38}$$

is the difference between J and its parallel transport in the direction tangent to ϕ, i.e. it is the value of the covariant derivative along ϕ. Since this difference is of order t^2, at $t = 0$ we have

$$\mathrm{D}_t^2 J = -K(J, \dot{\phi}), \tag{4.39}$$

or dropping the assumption that $\check{R}(e_1, e_2)\vec{e}_2$ is parallel to e_1,

$$\frac{\mathrm{D}^2 J}{\mathrm{d}t^2} + \check{R}(J, \dot{\phi})\vec{\dot{\phi}} = 0. \tag{4.40}$$

Considered as an equation for all $J(t)$, this is called the **Jacobi equation**, with the vector field $J(t)$ that satisfies it called a **Jacobi field**. A more precise way to generalize our construction of J is to define a one-parameter family of geodesics $\phi_s(t)$, so that

$$J(t) = \left.\frac{\partial \phi_s(t)}{\partial s}\right|_{s=0}. \tag{4.41}$$

If M is complete then every Jacobi field can be expressed in this way for some family of geodesics.

If we then consider the Jacobi fields $J_v(t)$ corresponding to the geodesics $\phi_v(t)$ of tangent vectors $\|v\| = 1$ parametrized by arclength and such that to order t we have $\|J_v(1)\| = 1$, it can be shown ([1] pp. 114-115) that to order t^3 we have $\|J_v(t)\| = t(1 - t^2 K(J_v, \dot{\phi}_v)/6)$.

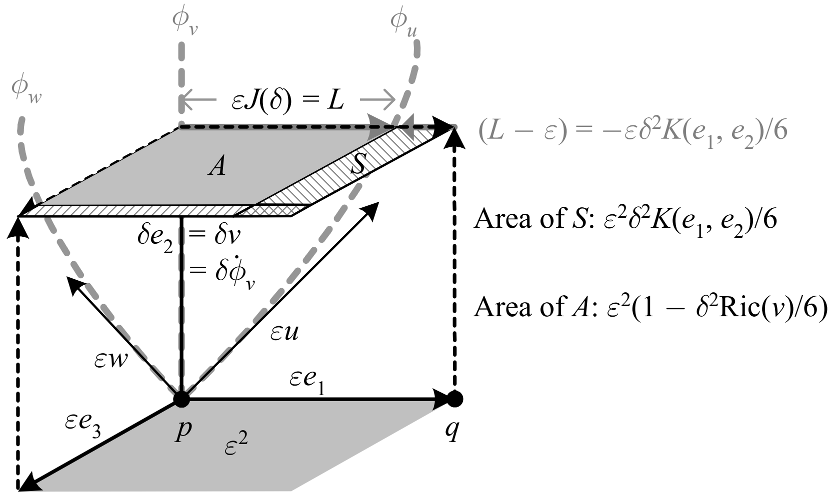

FIGURE 4.3: The infinitesimal geodesic area element derived from the Jacobi field between radial geodesics.

This means that if we apply the previous reasoning for parallel geodesics to these radial geodesics we have an "infinitesimal geodesic $(n-1)$-area element" $A(t) = t^2(1 - t^2\mathrm{Ric}(v)/6)$. Integrating this over all values of v gives for small $t = \varepsilon$ the surface area of a geodesic n-ball of radius ε, which we denote $\partial B_\varepsilon(M^n)$. But this integral just averages the values of the Ricci function, which is the Ricci scalar over the dimension n, so that to order ε^2 we have

$$\frac{\partial B_\varepsilon(M^n)}{\partial B_\varepsilon(\mathbb{R}^n)} = 1 - \frac{\varepsilon^2}{6n}R, \tag{4.42}$$

and integrating over the radius we find (see [5]) a similar relation for the volume of a geodesic sphere compared to a Euclidean one of

$$\frac{B_\varepsilon(M^n)}{B_\varepsilon(\mathbb{R}^n)} = 1 - \frac{\varepsilon^2}{6(n+2)}R. \tag{4.43}$$

Thus $\varepsilon^2 R/6n$ is "the fraction by which the surface area of a geodesic n-ball of radius ε is smaller than it would be under a flat metric," and $\varepsilon^2 R/6(n+2)$ is "the fraction by which the volume of a geodesic n-ball of radius ε is smaller than it would be under a flat metric."

Alternatively, we can use Riemann normal coordinates to express v in our "infinitesimal geodesic $(n-1)$-area element," whereupon following similar logic to the above we find that, at points close to the origin of our coordinates, to order $\|x\|^2$ the volume element is

$$\mathrm{d}V = \left(1 - \frac{1}{6}R_{\mu\nu}x^\mu x^\nu\right)\mathrm{d}x^1 \cdots \mathrm{d}x^n, \tag{4.44}$$

or using the expression of the volume element in terms the square root of the determinant of the metric, again to order $\|x\|^2$ we find that

$$g_{\mu\nu} = \delta_{\mu\nu} - \frac{1}{3}R_{\mu\lambda\nu\sigma}x^\lambda x^\sigma. \tag{4.45}$$

As is apparent from their definitions, the Ricci tensor and function do not depend on the metric. We can attempt to find a metric-free geometric interpretation by considering the concept of a **parallel volume form**. This is defined as a volume form which is invariant under parallel transport. We immediately see that it is only possible to define such a form if parallel transport around a loop does not alter volumes, i.e. that $\check{R}$ must be $o(r, s)$-valued. This means that the connection is metric compatible, so we can define one if we wish; but if we do not, and assume zero torsion so that the Ricci tensor is symmetric, then our logic for volumes remains valid and we can still take a metric-free view of the expression for $\mathrm{d}V$ above as expressing the geodesic volume as measured by the parallel volume form. Note that unlike the Ricci tensor and function, the definitions here of the individual sectional curvatures and scalar curvature do depend upon the metric.

4.8 Summary

Below, we review the intuitive meanings of the various relations we have defined on a Riemannian manifold.

Relation	Meaning
$\mathrm{div}(u)\mathrm{d}V = L_u \mathrm{d}V$	$\mathrm{div}(u)$ is the fraction by which a unit volume changes when transported by the flow of u.
$\int_V \mathrm{div}(u)\mathrm{d}V = \int_{\partial V} i_u \mathrm{d}V = \int_{\partial V} \langle u, n \rangle \, \mathrm{d}S$	The change in a volume due to transport by the flow of u is equal to the net flow of u across that volume's boundary.
$\mathrm{div}(u) = 0$	u having zero divergence means the flow of u leaves volumes unchanged, or the net flow of u across the boundary of a volume is zero.
$j \equiv \rho u$, ρ is the density of Q	The current vector j is the vector whose length is the amount of Q per unit time crossing a unit area perpendicular to j
$\frac{\mathrm{d}q}{\mathrm{d}t} = \Sigma - \int_{\partial V} \langle j, n \rangle \, \mathrm{d}S$	The change in q (the amount of Q within V) equals the amount generated less the amount which passes through ∂V.
$\frac{\partial \rho}{\partial t} = \sigma - \mathrm{div}(j)$	The change in the density of Q at a point equals the amount generated less the amount that moves away.

Table 4.4: Divergence and continuity relations and their intuitive meanings.

Relation	Meaning
$R \equiv g^{ab}R_{ab}$	The Ricci scalar is n times the average of the Ricci function on the set of unit tangent vectors.
$\text{Ric}(e_\mu) = \sum_{i \neq \mu} g_{\mu\mu} K(e_i, e_\mu)$	The Ricci function of a unit vector is $(n-1)$ times the average of the sectional curvatures of the planes that include the vector.
$R = \sum_j g_{jj} \text{Ric}(e_j)$	The Ricci scalar is n times the average of all the Ricci functions.
$R = 2 \sum_{i<j} K(e_i, e_j)$	The Ricci scalar is $n(n-1)$ times the average of all sectional curvatures.
$G(e_\mu, e_\mu) = -g_{\mu\mu} \sum_{\substack{i<j \\ i,j \neq \mu}} K(e_i, e_j)$	The Einstein tensor applied to a unit vector twice is $-(n-1)(n-2)/2$ times the average of the sectional curvatures of the planes orthogonal to the vector.
$d\left(\exp(\delta \hat{w}), \exp(\delta \parallel_{\varepsilon \hat{v}} \hat{w})\right) = \varepsilon \left(1 - \frac{\delta^2}{2} K(\hat{v}, \hat{w})\right)$	$K(\hat{v}, \hat{w})/2$ is the fraction by which the geodesic parallel to $\hat{w}$ starting $\hat{v}$ away bends towards $\hat{w}$.
$\ddot{L}(t) = -L(t) K(\hat{v}, \hat{w})$	$K(\hat{v}, \hat{w})$ is the acceleration of two parallel geodesics in the $\hat{w}$ direction with initial separation $\hat{v}$ towards each other as a fraction of the initial gap.
$\ddot{A}(t) = -A(t) \text{Ric}(v)$	$\text{Ric}(v)/2$ is the fraction by which the area defined by the geodesics emanating from the $(n-1)$-surface perpendicular to v changes in the direction of v. $\text{Ric}(v)$ is the acceleration of the parallel geodesics emanating from the $(n-1)$-surface perpendicular to v towards each other as a fraction of the initial surface.
$\frac{\partial B_\varepsilon(M^n)}{\partial B_\varepsilon(\mathbb{R}^n)} = 1 - \frac{\varepsilon^2}{6n} R$	$\varepsilon^2 R/6n$ is the fraction by which the surface area of a geodesic n-ball of radius ε is smaller than it would be under a flat metric.
$\frac{B_\varepsilon(M^n)}{B_\varepsilon(\mathbb{R}^n)} = 1 - \frac{\varepsilon^2}{6(n+2)} R$	$\varepsilon^2 R/6(n+2)$ is the fraction by which the volume of a geodesic n-ball of radius ε is smaller than it would be under a flat metric.

TABLE 4.5: Relations defined on a Riemannian manifold M^n and their intuitive meanings.

Appendices

A Tensors and forms

It is assumed the reader is familiar with vector spaces and inner products, as well as the tensor product and the exterior product (AKA wedge product, Grassmann product). In the following, we will limit our discussion to finite-dimensional real vector spaces $V = \mathbb{R}^n$; generalization to complex scalars is straightforward.

A.1 The structure of the dual space

Given a finite-dimensional vector space V, the **dual space** V^* is defined to be the set of linear mappings from V to the scalars (AKA the linear functionals on V). The elements of V^* can be added together and multiplied by scalars, so V^* is also a vector space, with the same dimension as V.

> △ Note that in general, the word "dual" is used for many concepts in mathematics; in particular, the dual space has no relation to the Hodge dual (defined below).

An element $\varphi\colon V \to \mathbb{R}$ of V^* is called a **1-form**. Given a pseudo inner product on V, we can construct an isomorphism between V and V^* defined by

$$v \mapsto \langle v, \ \rangle \,, \tag{A.1}$$

i.e. $v \in V$ is mapped to the element of V^* which maps any vector $w \in V$ to $\langle v, w\rangle$. This isomorphism then induces a corresponding pseudo inner product on V^* defined by

$$\langle \langle v, \ \rangle \,, \langle w, \ \rangle \rangle \equiv \langle v, w\rangle \,. \tag{A.2}$$

An equivalent way to set up this isomorphism is to choose a basis e_μ of V, and then form the **dual basis** β^ν of V^*, defined to satisfy $\beta^\lambda(e_\mu) = \delta^\lambda{}_\mu$. The isomorphism between V and V^* is then defined by the correspondence

$$v = v^\mu e_\mu \mapsto (\eta_{\mu\lambda} v^\mu)\beta^\lambda, \tag{A.3}$$

corresponding to the isomorphism induced by the pseudo inner product on V that makes e_μ orthonormal. Here we have used the **Einstein summation convention**, i.e. a repeated index implies summation. Note that if $\langle e_\mu, e_\mu\rangle = -1$ then $e_\mu \mapsto -\beta^\mu$. This isomorphism and its inverse (usually in the context of Riemannian manifolds) are called the **musical isomorphisms**, where if $v = v^\mu e_\mu$ and $\varphi = \varphi_\mu \beta^\mu$ we write

$$\begin{aligned} v^\flat &\equiv \langle v, \ \rangle \\ &= \left(\eta_{\mu\lambda} v^\lambda\right) \beta^\mu \\ &= v_\mu \beta^\mu \\ \varphi^\sharp &\equiv \langle \varphi, \ \rangle \\ &= \left(\eta^{\mu\lambda} \varphi_\lambda\right) e_\mu \\ &= \varphi^\mu e_\mu \end{aligned} \tag{A.4}$$

and call the $v^\flat$ the **flat** of the vector v and $\varphi^\sharp$ the **sharp** of the 1-form φ.

> △ It is important to remember that when the inner product is not positive definite, the signs of components may change under these isomorphisms. If the components are in terms of an arbitrary (non-orthonormal) basis, then as we will see in Section A.4, the components change their values as well, since $\eta_{\lambda\mu}$ is replaced by the metric tensor in the above analysis.

Note that since $\beta^\lambda(e_\mu) = \delta^\lambda{}_\mu$ and $\langle e_\mu, e_\lambda \rangle = \eta_{\mu\lambda}$ we have

$$\begin{aligned} \varphi(v) &= \varphi_\lambda \beta^\lambda(v^\mu e_\mu) \\ &= \varphi_\mu v^\mu \\ &= \eta_{\mu\lambda}\varphi^\lambda v^\mu \\ &= \left\langle \varphi^\sharp, v \right\rangle. \end{aligned} \tag{A.5}$$

☼ A 1-form acting on a vector can thus be viewed as yielding a projection. Specifically, with a positive definite inner product, $\varphi(v)/\left\|\varphi^\sharp\right\|$ is the length of the projection of v onto the ray defined by $\varphi^\sharp$.

It is important to note that there is no **canonical isomorphism** between V and V^*, i.e. we cannot uniquely associate a 1-form with a given vector without introducing extra structure, namely an inner product or a preferred basis. Either structure will do: a choice of basis is equivalent to the definition of the unique inner product on V that makes this basis orthonormal, which then induces the same isomorphism as that induced by the dual basis.

In contrast, a canonical isomorphism $V \cong V^{**}$ can be made via the association $v \in V \leftrightarrow \xi \in V^{**}$ with $\xi\colon V^* \to \mathbb{R}$ defined by $\xi(\varphi) \equiv \varphi(v)$. Thus V and V^{**} can be completely identified (for a finite-dimensional vector space), and we can view V as the dual of V^*, with vectors regarded as linear mappings on 1-forms.

Vector components are often viewed as a column vector, which means that 1-forms act on vector components as row vectors (which then are acted on by matrices from the right). Under a change of basis we then have the following relationships:

	Index notation	Matrix notation
Basis	$e'_\mu = A^\lambda{}_\mu e_\lambda$	$e' = eA$
Dual basis	$\beta'^\mu = (A^{-1})^\mu{}_\lambda \beta^\lambda$	$\beta' = A^{-1}\beta$
Vector components	$v'^\mu = (A^{-1})^\mu{}_\lambda v^\lambda$	$v' = A^{-1}v$
1-form components	$\varphi'_\mu = A^\lambda{}_\mu \varphi_\lambda$	$\varphi' = \varphi A$

Table A.1: Transformations under a change of basis.

Notes: A 1-form will sometimes be viewed as a column vector, i.e. as the transpose of the row vector (which is the sharp of the 1-form under a Riemannian signature). Then we have $(\varphi')^T = (\varphi A)^T = A^T \varphi^T$.

A.2 Tensors

A **tensor** of **type** (AKA valence) (m, n) is defined to be an element of the **tensor space**

$$V_{m,n} \equiv (V \otimes \cdots (m \text{ times}) \cdots \otimes V) \otimes (V^* \otimes \cdots (n \text{ times}) \cdots \otimes V^*). \tag{A.6}$$

A **pure tensor** (AKA simple or decomposable tensor) of type (m, n) is one that can be written as the tensor product of m vectors and n 1-forms; thus a general tensor is a sum of pure tensors. The integer $(m + n)$ is called the **order** (AKA degree, rank) of the tensor, while the tensor **dimension** is that of V. Vectors and 1-forms are then tensors of type $(1, 0)$ and $(0, 1)$. The

rank (sometimes used to refer to the order) of a tensor is the minimum number of pure tensors required to express it as a sum. In "tensor language" vectors $v \in V$ are called **contravariant vectors** and 1-forms $\varphi \in V^*$ are called **covariant vectors** (AKA covectors). A tensor of type $(k, 0)$ is then called a **contravariant tensor**, with **covariant tensors** being of type $(0, k)$, and other tensor types being called **mixed tensors**. Scalars can be considered tensors of type $(0, 0)$.

△ As noted above, the meanings of tensor rank and order are often swapped in the literature. Another potential source of confusion is that a mixed tensor is not the opposite of a pure tensor.

A.3 Tensors as multilinear mappings

There is an obvious multiplication of two 1-forms: the scalar multiplication of their values. The resulting object $\varphi\psi\colon V \times V \to \mathbb{R}$ is a nondegenerate bilinear form on V. Viewed as an "outer product" on V^*, multiplication is trivially seen to be a bilinear operation, i.e. $a\,(\varphi + \psi)\,\xi = a\varphi\xi + a\psi\xi$. Thus the product of two 1-forms is isomorphic to their tensor product.

We can extend this to any tensor by viewing vectors as linear mappings on 1-forms, and forming the isomorphism

$$\bigotimes \varphi_i \mapsto \prod \varphi_i. \tag{A.7}$$

Note that this isomorphism is not unique, since for example any real multiple of the product would yield a multilinear form as well. However it is canonical, since the choice does not impose any additional structure, and is also consistent with considering scalars as tensors of type $(0, 0)$.

We can thus consider tensors to be multilinear mappings on V^* and V. In fact, we can view a tensor of type (m, n) as a mapping from $i < m$ 1-forms and $j < n$ vectors to the remaining $(m - i)$ vectors and $(n - j)$ 1-forms.

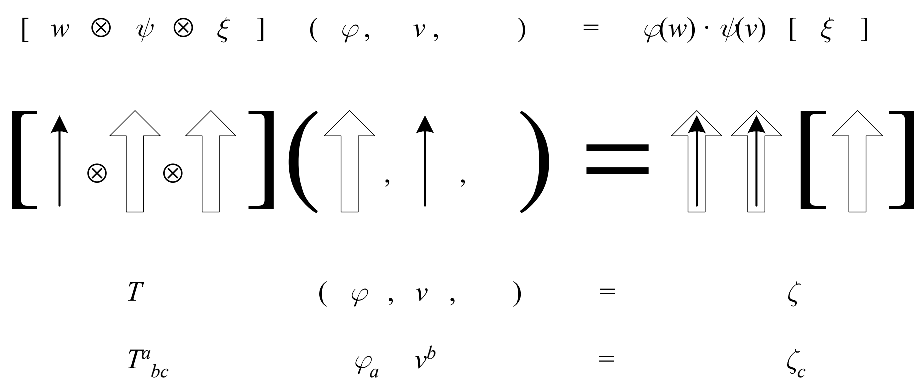

FIGURE A.1: Different ways of depicting a pure tensor of type $(1, 2)$. The first line explicitly shows the tensor as a mapping from a 1-form φ and a vector v to a 1-form ξ. The second line visualizes vectors as arrows, and 1-forms as receptacles that when matched to an arrow yield a scalar. The third line combines the constituent vectors and 1-forms of the tensor into a single symbol T while merging the scalars into ξ to define ζ, and the last line adds indices (covered in the next section).

A general tensor is a sum of pure tensors, so for example a tensor of the form $(u \otimes \varphi) + (v \otimes \psi)$ can be viewed as a linear mapping that takes ξ and w to the scalar $\xi\,(u) \cdot \varphi\,(w) + \xi\,(v) \cdot \psi\,(w)$.

Since the roles of mappings and arguments can be reversed, we can simplify things further by viewing the arguments of a tensor as another tensor:

$$\begin{aligned}(u \otimes \varphi)(\xi \otimes w) &\equiv (u \otimes \varphi)(\xi, w) \\ &= (\xi \otimes w)(u, \varphi) \\ &= \xi(u) \cdot \varphi(w)\end{aligned} \tag{A.8}$$

A.4 Abstract index notation

Abstract index notation uses an upper Latin index to represent each contravariant vector component of a tensor, and a lower index to represent each covariant vector (1-form) component. We can see from the preceding figure that this notation is quite compact and clearly indicates the type of each tensor while re-using letters to indicate what "slots" are to be used in the mapping.

The tensor product of two tensors ${S^a}_b \otimes {T^c}_d$ is simply denoted ${S^a}_b {T^c}_d$, and in this form the operation is sometimes called the **tensor direct product**. We may also consider a **contraction**

$${T^{ab}}_{bc} = {T^a}_c, \tag{A.9}$$

where two of the components of a tensor operate on each other to create a new tensor with a reduced number of indices. For example, if ${T^{ab}}_c = v^a \otimes w^b \otimes \varphi_c$, then ${T^{ab}}_b = \varphi(w) \cdot v^a$.

A (pseudo) inner product on V is a symmetric bilinear mapping, and thus corresponds to a symmetric tensor g_{ab} called the **(pseudo) metric tensor**. The isomorphism $v \in V \mapsto v^\flat \in V^*$ induced by this pseudo inner product is then defined by

$$v^a \mapsto v_a \equiv g_{ab} v^b, \tag{A.10}$$

and is called **index lowering**. The corresponding pseudo inner product on V^* is denoted g^{ab}, which provides a consistent **index raising** operation since we immediately obtain $g^{ab} g_{ac} g_{bd} = g_{cd}$. We also have the relation $v^a = g^{ab} v_b = g^{ab} g_{bc} v^c \Rightarrow g^{ab} g_{bc} = {g^a}_c = {\delta^a}_c$, the identity mapping. The inner product of two tensors of the same type is then the contraction of their tensor direct product after index lowering/raising, e.g. $\langle T^{ab}, S^{cd} \rangle = T^{ab} S_{ab} = T^{ab} g_{ac} g_{bd} S^{cd}$.

> △ It is important to remember that if v is a vector, the operation $v_a v^a$ implies index lowering, which requires an inner product. In contrast, if φ is a 1-form, the operation $\varphi_a v^a$ is always valid regardless of the presence of an inner product.

A symmetric or anti-symmetric tensor can be formed from a general tensor by adding or subtracting versions with permuted indices. For example, the combination $(T_{ab} + T_{ba})/2$ is the **symmetrized tensor** of T, i.e. exchanging any two indices leaves it invariant. The **anti-symmetrized tensor** $(T_{ab} - T_{ba})/2$ changes sign upon the exchange of any two indices, and yields the original tensor T_{ab} when added to the symmetrized tensor. The following notation is common for tensors with n indices, with the sums over all permutations of indices:

$$\text{Symmetrization: } T_{(a_1 \ldots a_n)} \equiv \frac{1}{n!} \sum_\pi T_{a_{\pi(1)} \ldots a_{\pi(n)}} \tag{A.11}$$

$$\text{Anti-symmetrization: } T_{[a_1 \ldots a_n]} \equiv \frac{1}{n!} \sum_{\pi} \operatorname{sign}(\pi)\, T_{a_{\pi(1)} \ldots a_{\pi(n)}} \tag{A.12}$$

This operation can be performed on any subset of indices in a tensor, as long as they are all covariant or all contravariant. Skipping indices is denoted with vertical bars, as in $T_{(a|b|c)} = (T_{abc} + T_{cba})/2$; however, note that vertical bars alone are sometimes used to denote a sum of ordered permutations, as in $T_{|abc|} = T_{abc} + T_{bca} + T_{cab}$.

A.5 Tensors as multi-dimensional arrays

In a given basis, a pure tensor of type (m, n) can be written using **component notation** in the form

$$v^1 \otimes \cdots \otimes v^m \otimes \varphi_1 \otimes \cdots \otimes \varphi_n \equiv T^{\mu_1 \ldots \mu_m}{}_{\nu_1 \ldots \nu_n} e_{\mu_1} \otimes \cdots \otimes e_{\mu_m} \otimes \beta^{\nu_1} \otimes \cdots \otimes \beta^{\nu_n}, \tag{A.13}$$

where the Einstein summation convention is used in the second expression. Note that the collection of terms into T is only possible due to the defining property of the tensor product being linear over addition. The tensor product between basis elements is often dropped in such expressions.

A general tensor is a sum of such pure tensor terms, so that any tensor T can be represented by a $(m+n)$-dimensional array of scalars. For example, any tensor of order 2 is a matrix, and type $(1, 1)$ tensors are linear mappings operating on vectors or forms via ordinary matrix multiplication if they are all expressed in terms of components in the same basis. Basis-independent quantities from linear algebra such as the trace and determinant are then well defined on such tensors.

△ A potentially confusing aspect of component notation is the basis vectors e_μ, which are not components of a 1-form but rather vectors, with μ a label, not an index. Similarly, the basis 1-forms β^ν should not be confused with components of a vector.

The Latin letters of abstract index notation (e.g. $T^{ab}{}_{cd}$) can thus be viewed as placeholders for what would be indices in a particular basis, while the Greek letters of component notation represent an actual array of scalars that depend on a specific basis. The reason for the different notations is to clearly distinguish tensor identities, true in any basis, from equations true only in a specific basis.

△ In general relativity both abstract and index notation are abused to represent objects that are non-tensorial (see Section 2).

△ Note that if abstract index notation is not being used, Latin and Greek indices are often used to make other distinctions, a common one being between indices ranging over three space indices and indices ranging over four space-time indices.

△ Note that "rank" and "dimension" are overloaded terms across these constructs: "rank" is sometimes used to refer to the order of the tensor, which is the dimensionality of the corresponding multi-dimensional array; the dimension of a tensor is that of the underlying vector space, and so is the length of a side of the corresponding array (also sometimes called the dimension of the array). However, the rank of a order 2 tensor coincides with the rank of the corresponding matrix.

A.6 Exterior forms as multilinear mappings

An **exterior form** (AKA k-form, alternating form) is defined to be an element of $\Lambda^k V^*$. Just as we formed the isomorphism $\otimes\varphi_i \mapsto \Pi\varphi_i$ to view tensors as multilinear mappings on V, we can view k-forms as alternating multilinear mappings on V. Restricting attention to the exterior product of k 1-forms $\bigwedge \varphi_i$, we define the isomorphism

$$\begin{aligned}\bigwedge_{i=1}^{k} \varphi_i &\mapsto \sum_{\pi} \operatorname{sign}(\pi) \prod_{i=1}^{k} \varphi_{\pi(i)} \\ &= \sum_{i_1, i_2, \ldots, i_k} \varepsilon^{i_1 i_2 \ldots i_k} \varphi_{i_1} \varphi_{i_2} \cdots \varphi_{i_k},\end{aligned} \tag{A.14}$$

where π is any permutation of the k indices, $\operatorname{sign}(\pi)$ is the sign of the permutation, and ε is the **permutation symbol** (AKA completely anti-symmetric symbol, Levi-Civita symbol, alternating symbol, ε-symbol), defined to be $+1$ for even index permutations, -1 for odd, and 0 otherwise.

☼ The above isomorphism extends the interpretation of forms acting on vectors as yielding a projection. Specifically, if the parallelepiped $\varphi^\sharp = \bigwedge \varphi_i^\sharp$ has volume V, then $\varphi(v_1, \ldots v_k)/V$ is the volume of the projection of the parallelepiped $v = \bigwedge v_i$ onto $\varphi^\sharp$.

Extending this to arbitrary forms $\varphi \in \Lambda^j V^*$ and $\psi \in \Lambda^k V^*$, we have

$$\begin{aligned}&(\varphi \wedge \psi)(v_1, \ldots, v_{j+k}) \\ &\mapsto \frac{1}{j!k!} \sum_{\pi} \operatorname{sign}(\pi)\, \varphi\left(v_{\pi(1)}, \ldots, v_{\pi(j)}\right) \psi\left(v_{\pi(j+1)}, \ldots, v_{\pi(j+k)}\right).\end{aligned} \tag{A.15}$$

Just as with tensors, this isomorphism is canonical but not unique; but in the case of exterior forms, other isomorphisms are in common use. The main alternative isomorphism inserts a term $1/k!$ in the first relation above, which results in $1/j!k!$ being replaced by $1/(j+k)!$ in the second. Note that this alternative is inconsistent with the interpretation of exterior products as parallelepipeds.

△ It is important to understand which convention a given author is using. The first convention above is common in physics, and we will adhere to it here.

A.7 Exterior forms as completely anti-symmetric tensors

An immediate result of this view of forms as multilinear mappings is that we can also view forms as completely anti-symmetric tensors under the identification of $\prod \varphi_i$ with $\bigotimes \varphi_i$. For example, for a 2-form we have the equivalent expressions

$$\begin{aligned}(\varphi \wedge \psi)(v, w) &\leftrightarrow (\varphi \otimes \psi - \psi \otimes \varphi)(v, w) \\ &\leftrightarrow \varphi(v)\, \psi(w) - \psi(v)\, \varphi(w)\,.\end{aligned} \tag{A.16}$$

Note however that this identification does not lead to equality of the inner products defined on tensors and exterior forms; instead for two k-forms we have

$$\left\langle \bigwedge \varphi_i, \bigwedge \psi_j \right\rangle_{\text{form}} = \det\left(\langle \varphi_i, \psi_j \rangle\right), \tag{A.17}$$

while as tensors we have

$$\left\langle \bigwedge \varphi_i, \bigwedge \psi_j \right\rangle_{\text{tensor}} = \left\langle \varepsilon^I \varphi_I, \varepsilon^J \varphi_J \right\rangle_{\text{tensor}} = k!\det\left(\langle \varphi_i, \psi_j \rangle\right). \tag{A.18}$$

Fortunately, the tensor inner product is almost always expressed explicitly in terms of index contractions, so we will continue to use the $\langle\ ,\ \rangle$ notation for the inner product of k-forms.

A.8 Exterior forms as anti-symmetric arrays

In terms of a basis β^μ of V^*, we can write a k-form φ as

$$\varphi = \frac{1}{k!} \sum_{\mu_1, \ldots, \mu_k} \varphi_{\mu_1 \ldots \mu_k} \beta^{\mu_1} \wedge \cdots \wedge \beta^{\mu_k}. \tag{A.19}$$

△ The above way of writing the components is not unique, and others are in common use, the main alternative omitting the factorial.

The advantage of the expression above is that, with our isomorphism convention, the component array can be identified with the anti-symmetric covariant tensor component array in the same basis:

$$\varphi \mapsto \frac{1}{k!} \varphi_{\mu_1 \ldots \mu_k} \sum_\pi \operatorname{sign}(\pi) \bigotimes_i \beta^{\pi(i)} = \varphi_{\mu_1 \ldots \mu_k} \beta^{\mu_1} \otimes \cdots \otimes \beta^{\mu_k} \tag{A.20}$$

Here we have dropped the summation sign in favor of the Einstein summation convention, and the last equality follows from the anti-symmetry of the component array.

A.9 Algebra-valued exterior forms

We can extend the view of exterior forms as real-valued linear mappings to define **algebra-valued forms**. These follow the same construction as in Section A.6 above, starting from an algebra-valued 1-form

$$\check{\Theta} \colon V \to \mathfrak{a}, \tag{A.21}$$

so that general forms are alternating multilinear maps from k vectors to a real algebra $\mathfrak{a}$ whose vector multiplication takes the place of multiplication in $\mathbb{R}$. Since this vector multiplication may not be commutative, we need to be more careful in terms of ordering in the isomorphism to ensure antisymmetry, i.e. for two algebra-valued 1-forms we define

$$(\check{\Theta} \wedge \check{\Psi})(v, w) \equiv \check{\Theta}(v)\check{\Psi}(w) - \check{\Theta}(w)\check{\Psi}(v). \tag{A.22}$$

An algebra-valued form whose values are defined by matrices is a **matrix-valued form**. Exterior forms that take values in a matrix group can also be considered as matrix-valued forms, but it must be understood that under addition the values may no longer be in the group. One can also form the exterior product between a matrix-valued form and a **vector-valued form**. To reduce confusion when dealing with algebra- and vector-valued forms, we will indicate them with (non-standard) decorations, for example in the case of a matrix-valued 1-form acting on a vector-valued 1-form,

$$(\check{\Theta} \wedge \vec{\varphi})(v, w) \equiv \check{\Theta}(v)\vec{\varphi}(w) - \check{\Theta}(w)\vec{\varphi}(v). \tag{A.23}$$

△ An additional distinction can be made between forms that take values which are concrete matrices and column vectors (and thus depend upon the basis of the underlying vector space), and forms that take values which are abstract linear transformations and abstract vectors (and thus are basis-independent). We will attempt to distinguish between these by referring to the specific matrix or abstract group, and by only using "vector-valued" when the value is an abstract vector.

A notational issue arises in the particular case of Lie algebra valued forms, where the related associative algebra in the relation $[\check{\Theta}, \check{\Psi}] = \check{\Theta}\check{\Psi} - \check{\Psi}\check{\Theta}$ could also be in use. In this case multiplication of values could use either the Lie commutator or that of the related associative algebra. We will denote the exterior product using the Lie commutator by $\check{\Theta}[\wedge]\check{\Psi}$. Some authors use $[\check{\Theta}, \check{\Psi}]$ or $[\check{\Theta} \wedge \check{\Psi}]$, but both can be ambiguous, motivating us to introduce our (non-standard) notation. The expression $\check{\Theta} \wedge \check{\Psi}$ is then reserved for the exterior product using the underlying associative algebra (e.g. that of matrix multiplication if the associative algebra is defined this way). For two Lie algebra-valued 1-forms we then have

$$\begin{aligned}(\check{\Theta}[\wedge]\check{\Psi})(v, w) &= [\check{\Theta}(v), \check{\Psi}(w)] - [\check{\Theta}(w), \check{\Psi}(v)] \\ &= \check{\Theta}(v)\check{\Psi}(w) - \check{\Psi}(w)\check{\Theta}(v) - \check{\Theta}(w)\check{\Psi}(v) + \check{\Psi}(v)\check{\Theta}(w).\end{aligned} \tag{A.24}$$

Note that $[\check{\Theta}, \check{\Psi}](v, w) = \check{\Theta}(v)\check{\Psi}(w) - \check{\Psi}(v)\check{\Theta}(w)$ is a distinct construction, as is $[\check{\Theta}(v), \check{\Psi}(w)] = \check{\Theta}(v)\check{\Psi}(w) - \check{\Psi}(w)\check{\Theta}(v)$; neither are in general anti-symmetric and thus do not yield forms. Also note that e.g. for two 1-forms $\check{\Theta}[\wedge]\check{\Psi} \neq \check{\Theta}\wedge\check{\Psi} - \check{\Psi}\wedge\check{\Theta}$, and $(\check{\Theta}[\wedge]\check{\Theta})(v, w) = 2[\check{\Theta}(v), \check{\Theta}(w)]$ does not in general vanish, so $[\wedge]$ does not act like a Lie commutator in these respects. However, for algebra-valued j- and k-forms $\check{\Theta}$ and $\check{\Psi}$, the operation $[\wedge]$ does in fact follow a graded commutativity rule

$$\check{\Theta}[\wedge]\check{\Psi} = (-1)^{jk+1}\,\check{\Psi}[\wedge]\check{\Theta}, \tag{A.25}$$

and with an algebra-valued m-form $\check{\Xi}$ we find a graded Jacobi identity of

$$(\check{\Theta}[\wedge]\check{\Psi})\,[\wedge]\,\check{\Xi} + (-1)^{j(k+m)}\,(\check{\Psi}[\wedge]\check{\Xi})\,[\wedge]\,\check{\Theta} + (-1)^{m(j+k)}\,(\check{\Xi}[\wedge]\check{\Theta})\,[\wedge]\,\check{\Psi} = 0. \tag{A.26}$$

Algebra-valued forms also introduce potentially ambiguous index notation. If a basis is chosen for the algebra $\mathfrak{a}$, the value of an algebra-valued form may be expressed using component notation as Θ^{μ}; or if the algebra is defined in terms of matrices, an element might be written $\Theta^{\alpha}{}_{\beta}$, an expression that has nothing to do with the basis of $\mathfrak{a}$. Then for example an algebra-valued 1-form might be written $\Theta^{\mu}{}_{\gamma}$ or $\Theta^{\alpha}{}_{\beta\gamma}$.

> △ In considering algebra-valued forms expressed in index notation, extra care must be taken to identify the type of form in question, and to match each index with the aspect of the object it was meant to represent.

A.10 The Hodge star

A pseudo inner product determines orthonormal bases for V, among which we can choose a specific one $\hat{e}_{\mu}$. The ordering of the $\hat{e}_{\mu}$ determines a choice of orientation. This orientation uniquely determines an orthonormal basis (i.e. a unit "length" vector) for the one-dimensional vector space $\Lambda^{n}V$, namely the **unit n-vector** (AKA orientation n-vector, volume element)

$$\Omega \equiv \hat{e}_1 \wedge \cdots \wedge \hat{e}_n. \tag{A.27}$$

> △ Many symbols are used in the literature for the unit n-vector and related quantities, including ε, i, I, and ω; to avoid confusion with the other common uses of these symbols, we will use the (non-standard) symbol Ω.

Since $\Lambda^{n}V$ is one-dimensional, every element of $\Lambda^{n}V$ is a real multiple of Ω. Thus Ω sets up a bijection (dependent upon the inner product and choice of orientation) between $\Lambda^{n}V$ and $\Lambda^{0}V = \mathbb{R}$. In general, $\Lambda^{k}V$ and $\Lambda^{n-k}V$ are vector spaces of equal dimension, and thus we can also set up a bijection between them.

The **Hodge star operator** (AKA Hodge dual) is defined to be the linear map

$$*\colon \Lambda^{k}V \to \Lambda^{n-k}V \tag{A.28}$$

that acts on $A \in \Lambda^{k}V$ such that for any $B \in \Lambda^{n-k}V$ we have

$$A \wedge B = \langle *A, B\rangle\,\Omega. \tag{A.29}$$

An equivalent requirement is that $\langle C \wedge *A, \Omega\rangle = \langle C, A\rangle$ for any $C \in \Lambda^{k}V$. In particular, for signature (r, s) we immediately obtain

$$A \wedge *A = \langle *A, *A\rangle\,\Omega = (-1)^{s}\,\langle A, A\rangle\,\Omega. \tag{A.30}$$

☼ These relations allow one to think of the Hodge star $*$ as an operator that roughly "swaps the exterior and inner products," or alternatively that yields the "orthogonal complement with the same magnitude."

The Hodge star operator is dependent upon a choice of inner product and orientation, but beyond that is independent of any particular basis. In particular, for any orthonormal basis $\hat{e}_\mu$ oriented with Ω, we can take $A \equiv \hat{e}_1 \wedge \cdots \wedge \hat{e}_k$ and $B \equiv \hat{e}_{k+1} \wedge \cdots \wedge \hat{e}_n$, in which case $*A = \langle B, B \rangle B$, i.e. $*A$ is constructed from an orthonormal basis for the orthogonal complement of A; in fact, this relation can be used as an equivalent definition of the Hodge star.

Below we list some easily derived facts about the Hodge star operator, where V is n-dimensional with unit n-vector Ω and a pseudo inner product of signature (r, s):

- $*\Omega = 1 \Rightarrow (*A)\,\Omega = A$ if $A \in \Lambda^n V$
- $*1 = (-1)^s\,\Omega \Rightarrow \langle \Omega, \Omega \rangle = (-1)^s$
- $**\,A = (-1)^{k(n-k)+s}\,A = (-1)^{k(n-1)+s}\,A$, where $A \in \Lambda^k V$
- $A \wedge *B = B \wedge *A = (-1)^s \langle A, B \rangle\,\Omega$ if $A, B \in \Lambda^k V$
- $\langle A, B \rangle = (-1)^s * (A \wedge *B)$ if $A, B \in \Lambda^k V$

△ Some authors instead define the Hodge star by the relation $A \wedge *B = \langle A, B \rangle\,\Omega$, which differs by a sign in some cases from the more common definition we use; in particular, with this definition $*\Omega = (-1)^s\Omega$ and $*1 = \Omega$.

Note that $*A$ is not a basis-independent object, since it reverses sign upon changing the chosen orientation. Such an object is prefixed by the word **pseudo-**, e.g. $*v$ is called a **pseudo-vector** (AKA axial vector, in which case v is called a polar vector) and Ω itself is a **pseudo-scalar**.

△ The use of "pseudo" to indicate a quantity that reverses sign upon a change of orientation should not be confused with the use of "pseudo" to indicate an inner product that is not positive-definite. There are also other uses of "pseudo" in use. In particular, in general relativity the term "pseudo-tensor" is sometimes used, where neither of the above meanings are implied; instead this signifies that the tensor is not in fact a tensor.

B Differentiable manifolds

Differentiable manifolds allow us to graft calculus onto a topological manifold, which we can think of as a "rubber sheet." The constructions of coordinates and tangent vectors enable us to define a family of derivatives associated with the concept of how vector fields change on the manifold. The challenge is in defining all these objects without an ambient space, which our intuitive picture normally depends upon.

△ Note that a differentiable manifold includes no concept of length or distance (a metric), and no structure that allows tangent vectors at different points to be compared or related to each other (a connection). It is important to remember that nothing in this section depends upon these two extra structures.

When dealing with manifolds, there are two main approaches one can take: express everything in terms of coordinates, or strive to express everything in a coordinate-free fashion. In keeping with our attempt to focus on concepts rather than calculations, we will take the latter approach, but will take pains to carefully express fundamental concepts in terms of coordinates in order to derive a picture of what these coordinate-free tools do.

B.1 Coordinates

A key feature of a topological manifold M^n is that every point has an open neighborhood homeomorphic to an open subset of $\mathbb{R}^n$. To make this precise we define the following terms.

- **Coordinate chart** (AKA parameterization, patch, system of coordinates): a homeomorphism $\alpha\colon U \to \mathbb{R}^n$ from an open set $U \subset M^n$ to an open subset of $\mathbb{R}^n$
- **Coordinate functions** (AKA coordinates): the maps $a^\mu\colon U \to \mathbb{R}$ that project α down to one of the canonical Cartesian components
- **Atlas**: a collection of coordinate charts that cover the manifold
- **Coordinate transformation** (AKA change of coordinates, transition function): in a region covered by two charts, we can construct the map $\alpha_2 \circ \alpha_1^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$

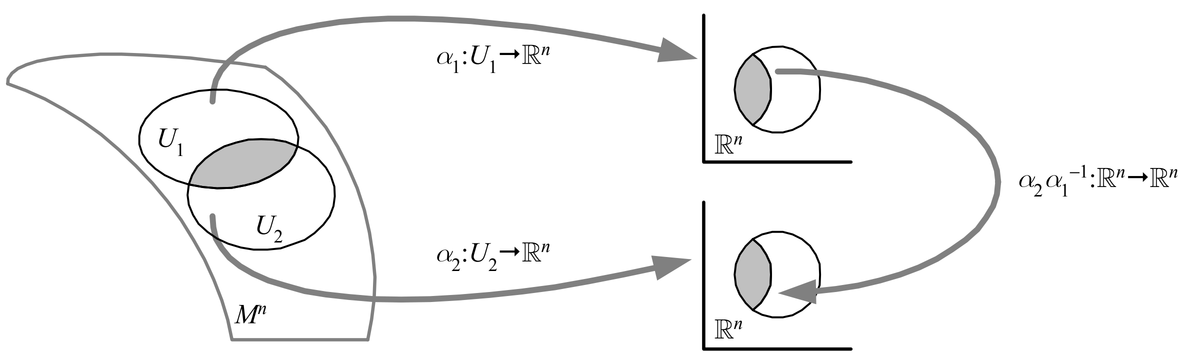

FIGURE B.1: In the intersection of two coordinate charts we can construct the coordinate transformation, a homeomorphism on $\mathbb{R}^n$.

△ A coordinate chart is sometimes defined to be the inverse map $\alpha^{-1}\colon \mathbb{R}^n \to M$ valid on an open subset of $\mathbb{R}^n$, with similar changes to related definitions such as coordinate functions.

The coordinate transformations are simply maps on Euclidean space, so we can require them to be infinitely differentiable (AKA smooth, C^∞). An atlas whose charts all have smooth coordinate transformations determines a **differentiable structure**, which turns the topological manifold into a **differentiable manifold** (AKA smooth manifold). Two differential structures are considered to be equivalent if the union of their atlases still results in smooth coordinate transformations. Unless otherwise noted, from this point forward "manifold" will mean differentiable manifold.

A **complex manifold** is defined to have an atlas of charts to $\mathbb{C}^n$ whose coordinate transformations are analytic. Complex n-manifolds are a subset of real $2n$-manifolds, but atlases are highly constrained since complex analytic functions are much more constrained than smooth functions. By "manifold" we will always mean a real manifold.

With the addition of a differentiable structure, one can define the various tools of calculus on manifolds in a straightforward way. Differentiable functions $f\colon U \to \mathbb{R}$ require the map $f \circ \alpha^{-1}\colon \mathbb{R}^n \to \mathbb{R}$ to be differentiable, and differentials $\partial/\partial a^\mu$ are defined at a point $p \in U$ by

$$\frac{\partial}{\partial a^\mu}(f)\bigg|_p \equiv \frac{\partial}{\partial x^\mu}\left(f \circ \alpha^{-1}(x)\right)\bigg|_{x=\alpha(p)}. \tag{B.1}$$

where $x \in \mathbb{R}^n$. All of the usual relations of calculus hold with these definitions.

△ To avoid clutter, a common abuse of notation is to use x^μ to denote any or all of three quantities: the point $p \in M$, the coordinate functions $a^\mu\colon M \to \mathbb{R}$, and the $\mathbb{R}^n$ n-tuplet $x^\mu = a^\mu(p)$. Similarly, the differential $\partial/\partial a^\mu$ is often denoted $\partial/\partial x^u$. We will follow these conventions going forward, but when dealing with fundamental definitions or pictures, it is important to distinguish these very different quantities from each other. Another shortcut is to denote differentials by ∂_μ; as with basis vectors, it is important to remember that these are labels, not component indices.

B.2 Tangent vectors and differential forms

The **tangent space** T_pU at a point $p \in U$ is defined to be the vector space spanned by the differential operators $\partial/\partial a^\mu\ |_p$. A **tangent vector** $v \in T_pU$ can then be expressed in tensor component notation as $v = v^\mu \partial/\partial a^\mu$, so that $v(a^\mu) = v^\mu$. The tangent vector $\partial/\partial a^\mu\ |_p$ applied to a function f can be thought of as "the change of f in the direction of the μ^{th} coordinate line at p."

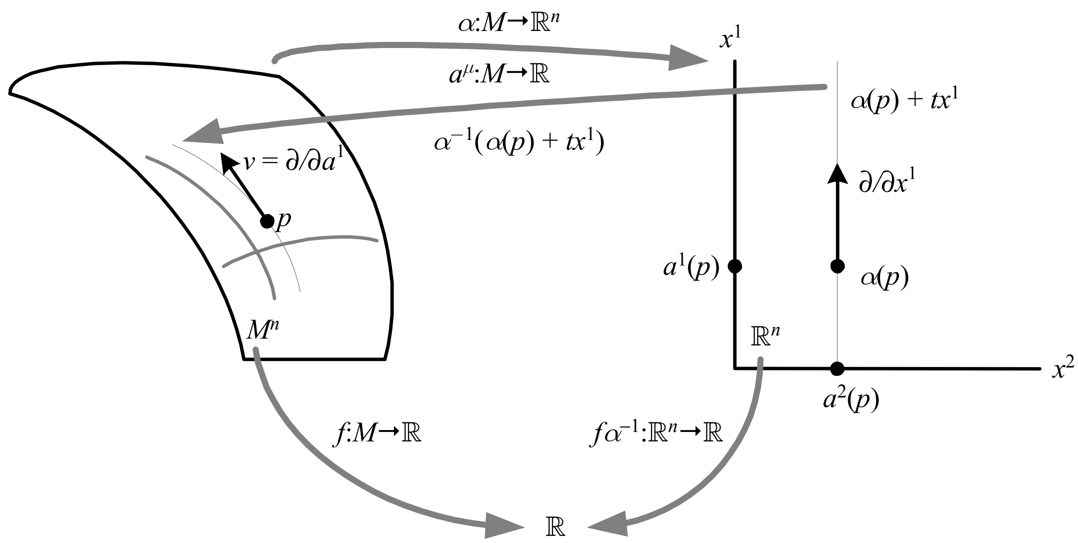

FIGURE B.2: In a particular coordinate chart, a tangent vector v operates on a function by taking the derivative of the composite function in $\mathbb{R}^n$ in the direction of $v^\mu \partial/\partial a^\mu$.

Thus at a point p, we have

$$v^\mu \frac{\partial}{\partial a^\mu}(f) = v^\mu \frac{\partial}{\partial x^\mu}\left(f \circ \alpha^{-1}(x)\right), \tag{B.2}$$

where $x = \alpha(p)$. The coordinate line $\alpha^{-1}\left(a^\mu(p) + tv^\mu x^\mu\right)$ is a parameterized curve on M, and thus it and the tangent vector itself are coordinate-independent objects. In another coordinate chart, the coordinate line that yields the same operator on functions near p can be seen to correspond to the familiar transformation of vector components

$$v = v^\mu \frac{\partial}{\partial a^\mu} = \left(v^\lambda \frac{\partial b^\mu}{\partial a^\lambda}\right)\frac{\partial}{\partial b^\mu}. \tag{B.3}$$

We can consider the point "p moved in the direction v" by abusing notation to write $p^\mu + tv^\mu$ in place of $\alpha^{-1}\left(a^\mu(p) + tv^\mu x^\mu\right)$; this is a coordinate-dependent expression, but in the limit $\varepsilon \to 0$ we can unambiguously write $p + \varepsilon v$ to refer to the concept "p moved infinitesimally in the direction v," which is coordinate-independent. This allows us to write

$$v(f) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[f_{p+\varepsilon v} - f_p\right]. \tag{B.4}$$

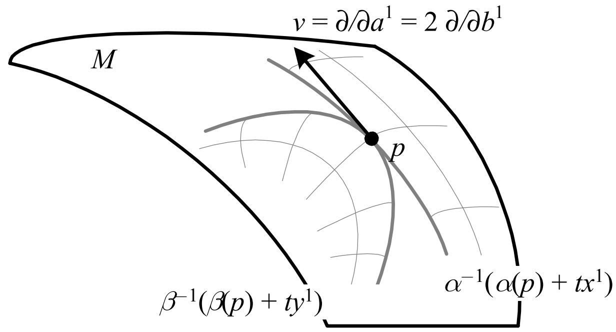

FIGURE B.3: A tangent vector v in terms of two different coordinate charts. $v^\mu = (1, 0)$ in chart α with coordinate functions $a^\mu(p) = x^\mu$, and $v^\mu = (2, 0)$ in chart β with coordinate functions $b^\mu(p) = y^\mu$. The divergent coordinate lines show that the concept of moving a point "in the direction of v" can only be coordinate-independent in the infinitesimal limit.

The set of all tangent spaces in a region U is called the **tangent bundle**, and is denoted TU. A (smooth, contravariant) **vector field** on U is then a tangent vector defined at each point such that its application to a smooth function on U is again smooth. Similarly, a **covariant vector field** is a 1-form defined at each point such that its value on a vector field is a smooth function, and a **tensor field** is the tensor product of vector fields and covariant vector fields.

△ Tensor fields (including vector fields and covariant vector fields) are written using the same notation as tensors, making it important to distinguish the two situations. In particular, one can define a (pseudo) metric tensor field, which is then usually referred to as simply a metric.

Note that a tensor field must remain a tensor locally at any point p, i.e. it must be a multilinear mapping. For example, a covariant tensor field can only depend upon the values of its vector field arguments at p, since otherwise one could add a vector field that vanishes at p and obtain a different result. This means that operators such as the derivatives on manifolds we will see in Sections C and 2 cannot usually be viewed as tensors, since they measure the difference between arguments at different points.

Since vectors are operators on functions, we can apply one vector field to another. Following the practice of using $\partial/\partial x^u$ to refer to $\partial/\partial a^\mu$, this can be used to define the **Lie bracket of vector fields**

$$\begin{aligned} [v, w](f) &\equiv v(w(f)) - w(v(f)) \\ \Rightarrow [v, w] &= \left(v^\mu \frac{\partial w^\lambda}{\partial x^\mu} - w^\mu \frac{\partial v^\lambda}{\partial x^\mu} \right) \frac{\partial}{\partial x^\lambda}. \end{aligned} \tag{B.5}$$

Here we have used the equality of mixed partials, and can easily verify that $[v, w]$ is anti-commuting and satisfies the Jacobi identity. Since this expression is coordinate-independent, $[v, w]$ is a vector field and we can thus view vect(M), the set of all vector fields on M, as the infinite-dimensional **Lie algebra of vector fields** on M, with vector multiplication defined by the Lie bracket.

Having defined vector and tensor fields on manifolds, we can now define a **differential form** as an alternating covariant tensor field, i.e. an exterior form in $\Lambda(T_pU)$ smoothly defined for every point p.

> △ Just as tensor fields are usually referred to as simply tensors, differential forms are usually referred to as simply **forms**, and a k-form is written simply $\varphi \in \Lambda^k M$. It is important to remember that in the context of manifolds, a k-form is an exterior form smoothly defined on k elements of the tangent space at each point, i.e. an anti-symmetric covariant k-tensor field.

On a differentiable manifold, the existence of k-forms makes possible a more concrete definition of orientability: a manifold M^n is orientable if there exists a non-vanishing n-form. Such a form is called a **volume form** (AKA volume element), since it gains a Jacobian-like determinant factor under invertible linear transformations.

> △ The term "volume form" or "volume element" is sometimes defined in physics to reflect the intuitive idea of a form which returns the volume spanned by its argument vectors; however, volume is always positive, so that in this usage we are more accurately referring to a **volume pseudo-form** whose value is the absolute value of the volume form as we have defined it.

B.3 Frames

A **frame** e_μ on $U \subset M^n$ is defined to be a tensor field of bases for the tangent spaces at each point, i.e. n linearly independent smooth vector fields e_μ.

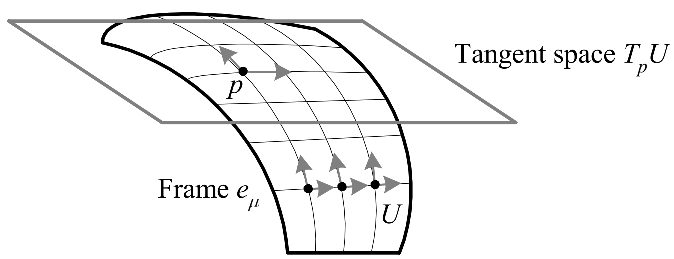

FIGURE B.4: A frame e_μ is n smooth vector fields that together provide a basis for the tangent space at every point.

The concept of frame has a particularly large number of synonyms, including comoving frame, repï¿½re mobile, vielbein, n-frame, and n-bein (where n is the dimension). The **dual frame**, the 1-forms β^μ corresponding to a frame e_μ, is also often simply called the frame.

When using particular coordinates x^μ, the frame $e_\mu = \partial/\partial x^\mu$ is called the **coordinate frame** (AKA coordinate basis or associated basis); any other frame is then called a **non-coordinate frame**. A **holonomic frame** is a coordinate frame in some coordinates (though perhaps not the ones being used); this condition is equivalent to requiring that $[e_\mu, e_\nu] = 0$, a result which is sometimes called **Frobenius' theorem**. An **anholonomic frame** is then a frame that cannot be derived from any coordinate chart in its region of definition. Using a non-coordinate frame suited to a specific problem is sometimes called the **method of moving frames**.

> △ Note that the distinction between holonomic and coordinate frames as defined here is often not made.

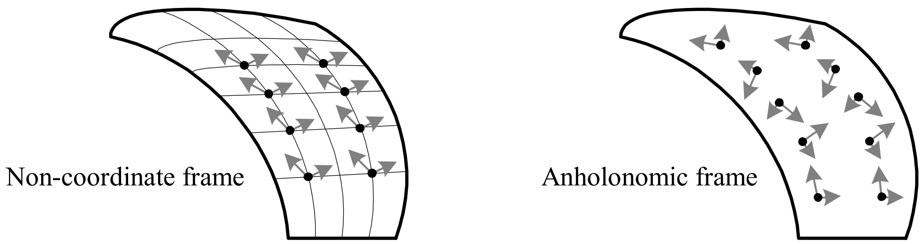

FIGURE B.5: A non-coordinate frame is not tangent to the coordinate functions being used, while an anholonomic frame cannot be derived from any coordinate chart.

A frame cannot usually be globally defined on a manifold. A simple way to see this is by the example of the 2-sphere S^2. Any drawing of coordinate functions on a globe will have singularities, such as the north and south poles when using latitude and longitude; these are points where the associated coordinate frame will either be undefined or will vanish. In general, there is no non-zero smooth vector field that can be defined on S^n for even n (this is sometimes called the **hedgehog theorem**, AKA hairy ball theorem, coconut theorem).

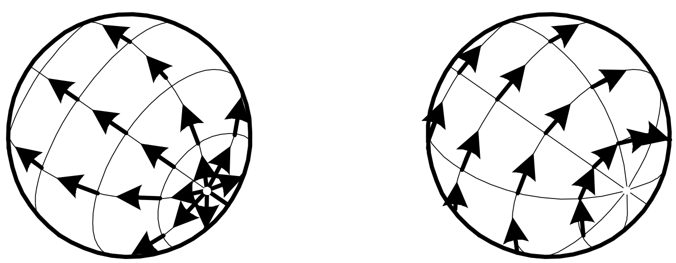

FIGURE B.6: The hedgehog theorem for S^2, showing that any attempt to "comb the hair of a hedgehog" yields bald spots, in this case at the poles.

A manifold that can have a global frame defined on it is called **parallelizable**. Some facts regarding parallelizable manifolds include:

- All parallelizable manifolds are orientable (and therefore have a volume form), but as we saw with S^2 the converse is not in general true
- Any orientable 3-manifold M^3 is parallelizable $\Rightarrow$ any 4-manifold $M^3 \times \mathbb{R}$ is parallelizable (important in the case of the spacetime manifold)
- Of the n-spheres, only S^1, S^3, and S^7 are parallelizable (this can be seen to be related to $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$ being the only normed finite-dimensional real division algebras beyond $\mathbb{R}$)

- The torus (with any number of holes) is the only closed orientable surface with a non-zero smooth vector field

B.4 Tangent vectors in terms of frames

It is important to remember that in following our intuitive picture of a Euclidean surface, our central definitions were manifolds M and tangent vectors v. These are the "real" intrinsic objects, while their expressions in terms of a particular coordinate chart and frame are arbitrary. Coordinates and frames are "temporary" tools we use to "componentize" points and tangents on a manifold.

In particular, if a manifold is defined in terms of a set of coordinate functions that feature a singularity, this singularity may be due to the coordinates extending outside of their valid chart, telling us nothing about whether the manifold itself has a singularity. Every point of a well-defined differentiable manifold always has a local coordinate chart and tangent vectors.

For example, given the typical spherical coordinate chart for S^2 the associated frame will be singular at the poles, since they are outside of U for that chart; nevertheless, tangent vectors are well-defined at these points, and can be expressed perfectly normally in a different chart.

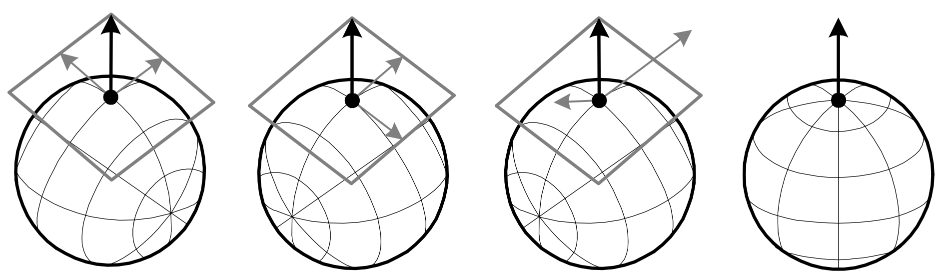

FIGURE B.7: A manifold and tangent vector expressed in terms of different coordinate functions and frames.

In the above figure, we see the following situations depicted:

- $v = e_1 + e_2 = \partial/\partial x^1 + \partial/\partial x^2$ (expressed in a coordinate frame)
- $v = e'_1 - e'_2 = \partial/\partial x'^1 - \partial/\partial x'^2$ (using a different coordinate frame)
- $v = e''_1 + 3e''_2 = \partial/\partial x'^1 - \partial/\partial x'^2$ (in a non-coordinate frame)

The final figure depicts coordinate functions that are singular at the point of interest; the manifold and vector are still well-defined, but the tangent space at this point cannot be expressed in terms of this coordinate chart.

△ In general, when working with objects on manifolds, it is important to keep clearly in mind whether a given symbol represents a vector, form, or function (0-form); whether any given index is a label, an abstract index or a component index in a particular frame or coordinates; and whether the object is a field with a value at each point, or is only valid at a particular point. Any calculation can always be made explicit by expressing everything in terms of functions and differential operators on them.

B.5 Diffeomorphisms

In the same way that spaces or topological manifolds are equivalent if they are related by a homeomorphism, differentiable manifolds are equivalent if they are related by a **diffeomorphism**, a homeomorphism that is differentiable along with its inverse. As usual we define differentiability by moving the mapping to $\mathbb{R}^n$, e.g. $\Phi\colon M \to N$ is differentiable if $\alpha_N \circ \Phi \circ \alpha_M^{-1}\colon \mathbb{R}^m \to \mathbb{R}^n$ is, where α_M and α_N are charts for M and N. Intuitively, a diffeomorphism like a homeomorphism can be thought of as arbitrary stretching and bending, but it is "nicer" in that it preserves the differentiable structure.

△ It is important to distinguish between coordinate transformations, which are locally defined and so may have singularities outside of a given region; and diffeomorphisms, which are globally defined and form a group. One can define a coordinate transformation on a region of a manifold that avoids any resulting singularities, but a diffeomorphism must be smooth on the entire manifold.

B.6 The differential and pullback

If we consider a general mapping between manifolds $\Phi\colon M^m \to N^n$, we can choose charts $\alpha_M\colon M \to \mathbb{R}^m$ and $\alpha_N\colon N \to \mathbb{R}^n$, with coordinate functions x^μ and y^ν, so that the mapping $\alpha_N \circ \Phi\colon M \to \mathbb{R}^n$ can be represented by n functions $\Phi^\nu\colon M \to \mathbb{R}$. This allows us to write down an expression for the induced **tangent mapping** or **differential** (aka pushforward, derivative) $\mathrm{d}\Phi\colon TM \to TN$ (also denoted $T\Phi$ or Φ_* or sometimes simply Φ if it is clear the argument is a tangent vector). For a tangent vector $v = v^\mu \partial/\partial x^\mu$ at a point $p \in M$ we define

$$\mathrm{d}\Phi\left(v\right)\big|_p \equiv v^\mu \frac{\partial \Phi^\nu}{\partial x^\mu}\frac{\partial}{\partial y^\nu}\bigg|_{\Phi(p)}. \tag{B.6}$$

This definition can be shown to be coordinate-independent and to follow our intuitive expectation that mapped tangent vectors stay tangent to mapped curves. If $M = N$ and Φ is the identity, $\mathrm{d}\Phi$ is just the vector component transformation in Section B.2. The matrix

$$J_\Phi(x) \equiv \partial\Phi^\nu/\partial x^\mu \tag{B.7}$$

is called the **Jacobian matrix** (AKA Jacobian).

If Φ is a diffeomorphism, $\mathrm{d}\Phi$ is an isomorphism between the tangent spaces at every point in M. The **inverse function theorem** says that the converse is true locally: if $\mathrm{d}\Phi_p$ is an isomorphism at $p \in M$, then Φ is locally a diffeomorphism. In particular, this means that if in some coordinates the Jacobian matrix is nonsingular, then $\alpha_N \circ \Phi \circ \alpha_M^{-1}$ represents a locally valid coordinate transformation and $\Phi^\nu = y^\nu$.

A mapping between manifolds $\Phi\colon M^m \to N^n$ also can be used to naturally define the **pullback** of a form $\Phi^*\colon \Lambda^k N \to \Lambda^k M$ by

$$\Phi^*\varphi\left(v_1, \ldots, v_k\right) = \varphi\left(\mathrm{d}\Phi\left(v_1\right), \ldots, \mathrm{d}\Phi\left(v_k\right)\right), \tag{B.8}$$

where the name indicates that a form on N can be "pulled back" to M using Φ. Note that the composition of pullbacks is then

$$\Psi^*\Phi^*\varphi = (\Phi\Psi)^*\varphi. \tag{B.9}$$

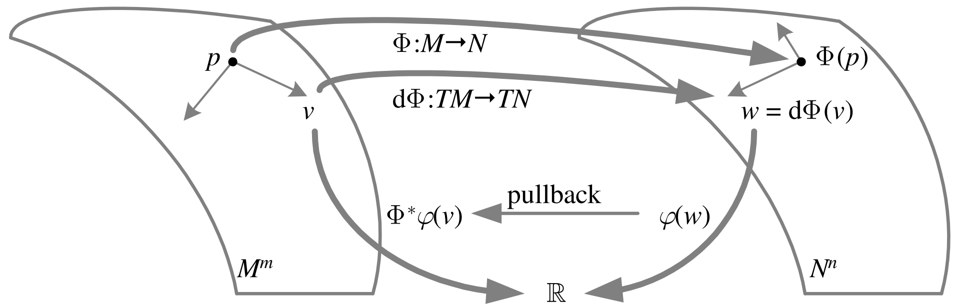

FIGURE B.8: Forms φ on N are pulled back to M by sending argument vectors to N using $\mathrm{d}\Phi$.

Note that for a mapping $f\colon M \to \mathbb{R}$, we have $\mathrm{d}f\colon TM \to T\mathbb{R} \cong \mathbb{R}$, so that $\mathrm{d}f\,(v) = v^\mu \partial f/\partial x^\mu = v\,(f)$, the directional derivative of f. Let us apply this to the coordinate function $x^1\colon M \to \mathbb{R}$. Then we have $\mathrm{d}x^1\,(v) = v^\mu \partial x^1/\partial x^\mu = v^1$, so that in particular $\mathrm{d}x^\nu\,(\partial/\partial x^\mu) = \delta^\nu{}_\mu$, i.e. $\mathrm{d}x^\mu$ is in fact the dual frame to $\partial/\partial x^\mu$. Thus in a given coordinate system, we can write a general tensor of type (m,n) as

$$T = T^{\mu_1 \dots \mu_m}{}_{\nu_1 \dots \nu_n} \frac{\partial}{\partial x^{\mu_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_m}} \otimes \mathrm{d}x^{\nu_1} \otimes \cdots \otimes \mathrm{d}x^{\nu_n}. \tag{B.10}$$

In particular, the metric tensor is often written

$$\mathrm{d}s^2 \equiv g = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{B.11}$$

where the Einstein summation convention is used and the tensor symbol omitted. A general k-form $\varphi \in \Lambda^k M$ can then be written as

$$\varphi = \sum_{\mu_1 < \cdots < \mu_k} \varphi_{\mu_1 \dots \mu_k} \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_k}. \tag{B.12}$$

From either the tangent mapping definition or the behavior of the exterior product under a change of basis, we can see that under a change of coordinates we have

$$\mathrm{d}y^{\mu_1} \wedge \cdots \wedge \mathrm{d}y^{\mu_k} = \det\left(\frac{\partial y^\nu}{\partial x^\mu}\right) \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_k}. \tag{B.13}$$

This is the familiar **Jacobian determinant** (like the Jacobian matrix, also often called the Jacobian) that appears in the change of coordinates rule for integrals from calculus, and explains the name of the volume form as defined previously in terms of the exterior product.

In summary, the differential d has a single definition, but is used in several different settings that are not related in an immediately obvious way.

Construct	Argument	Other names	Other symbols
$\mathrm{d}\Phi\colon TM \to TN$	$\Phi\colon M \to N$	Tangent mapping	$T\Phi$, Φ_*, Φ
$\mathrm{d}f\colon TM \to \mathbb{R}$	$f\colon M \to \mathbb{R}$	Directional derivative	$v\,(f)$, $\mathrm{d}_v f$, $\nabla_v f$
$\mathrm{d}x^\mu\colon TM \to \mathbb{R}$	$x^\mu\colon M \to \mathbb{R}$	Dual frame to $\partial/\partial x^\mu$	β^μ

TABLE B.1: Various uses of the differential on manifolds.

B.7 Immersions and embeddings

We can generalize and make precise the concept of a surface embedded in 3-dimensional space with the following definitions concerning a differentiable map $\Phi\colon M^m \to N^n$:

- **Immersion**: $\mathrm{d}\Phi$ is injective for all $p \in M$; intuitively, a smooth mapping that doesn't collapse the tangent spaces
- **Submanifold**: an immersion with Φ injective; intuitively, an immersion that doesn't intersect itself
- **Embedding** (AKA imbedding): a submanifold with Φ a homeomorphism onto $\Phi(M)$; intuitively, a submanifold that doesn't have intersecting limit points

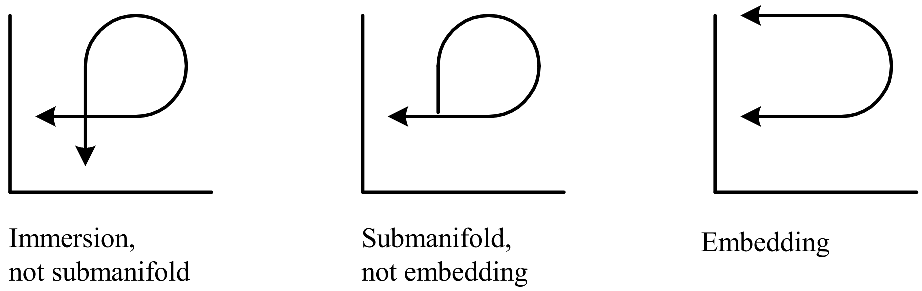

FIGURE B.9: $\mathbb{R}$ immersed in $\mathbb{R}^2$; the second immersion approaches a self-intersection in the limit as the line approaches infinity.

The difference in dimension $(n-m)$ is called the **codimension** of the embedding. The **Whitney embedding theorem** states that for positive codimension, any M^m can be immersed in $\mathbb{R}^{(2m-1)}$ and embedded in $\mathbb{R}^{2m}$. Thus we can view differentiable manifolds as generalized surfaces that we study without making reference to the enclosing Euclidean space. The limiting dimension of this theorem is illustrated by noting that the real projective space $\mathbb{R}\mathrm{P}^m$ cannot be embedded in $\mathbb{R}^{(2m-1)}$.

C Derivatives on manifolds

In this section we will introduce various objects that in some way measure how vectors or forms change from point to point on a manifold.

C.1 Derivations

In general, we define a **derivation** to be a linear map $\mathcal{D}\colon \mathfrak{a} \to \mathfrak{a}$ on an algebra $\mathfrak{a}$ that follows the **Leibniz rule** (AKA product rule)

$$\mathcal{D}(AB) = (\mathcal{D}A)B + A(\mathcal{D}B). \tag{C.1}$$

As noted previously in Section B.2, the set $\mathrm{vect}(M)$ of vector fields on a manifold form a Lie algebra; the Lie bracket operation with a fixed vector field $[u,\ ]$ is then a derivation on this algebra, since the Leibniz rule

$$[u,[v,w]] = [[u,v],w] + [v,[u,w]] \tag{C.2}$$

is just the Jacobi identity.

For a graded algebra, e.g. the exterior algebra, the **degree** of a derivation is the integer c where $\mathcal{D}\colon \Lambda^k M \to \Lambda^{k+c} M$. A **graded derivation** is defined to follow the **graded Leibniz rule**, e.g. for a k-form φ,

$$\mathcal{D}(\varphi \wedge \psi) = \mathcal{D}\varphi \wedge \psi + (-1)^{kc}\, \varphi \wedge \mathcal{D}\psi. \tag{C.3}$$

If c is odd, a graded derivation is sometimes called an **anti-derivation** (AKA skew-derivation).

C.2 The Lie derivative of a vector field

Without some kind of additional structure, there is no way to "transport" vectors, or compare them at different points on a manifold, and therefore no way to construct a vector derivative. The simplest way to introduce this structure is via another vector field, which leads us to the **Lie derivative**

$$L_v w \equiv [v,w]. \tag{C.4}$$

As noted above, L_v is a derivation due to the Jacobi identity. In this section we define the Lie derivative in terms of infinitesimal vector transport, and explore its geometrical meaning.

Given any vector field v on M, it can be shown ([2] pp. 125-127) that there exists a parameterized curve $v_p(t)$ at every point $p \in M$ such that $v_p(0) = p$ and $\dot{v}_p(t)$ is the value of the vector field v at the point $v_p(t)$ (the dot indicates the derivative with respect to t, which as usual is calculated on the curve mapped to $\mathbb{R}^n$ by the coordinate chart). Each curve in this family is in general only well-defined locally, i.e. for $-\varepsilon < t < \varepsilon$, and is thus called the **local flow** of v.

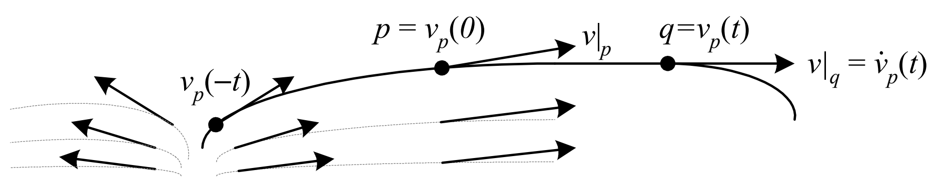

FIGURE C.1: A depiction of the local flow of a vector field v, with details on the local parameterized curve $v_p(t)$ at a point p.

For a fixed value of t, there is some region $U \subset M$ where the map $\Phi_t\colon U \to U$ defined by $p \mapsto v_p(t)$ is a diffeomorphism, and within the valid domain of t the maps Φ_t satisfy the abelian group law $\Phi_t \circ \Phi_s = \Phi_{t+s}$; thus the Φ_t are called a **local one-parameter group of diffeomorphisms**. This name is somewhat misleading, since due to the limited domain of t the maps Φ_t do not actually form a group; the "local" reflects the fact that the diffeomorphisms are not on all of M. In the case that these maps are in fact valid for all of t and M, v is called a **complete vector field**, and the Φ_t are called a **one-parameter group of diffeomorphisms**. If M is compact, then every vector field is complete; if not, then a vector field is complete if it has **compact support** (is non-zero on a compact subset of M).

The tangent map dΦ defined by the vector field v is then the extra structure we need to "transport" vectors. dΦ maps a vector tangent to the curve C to a vector tangent to the curve $\Phi(C)$; it "pushes vectors along the flow of v." We can now define the Lie derivative as a limit

$$\begin{aligned} L_v w &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[\mathrm{d}\Phi_{-\varepsilon}\left(w\,|_{v_p(\varepsilon)}\right) - w\,|_p\right] \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[w\,|_{v_p(\varepsilon)} - \mathrm{d}\Phi_{\varepsilon}\left(w\,|_p\right)\right]. \end{aligned} \tag{C.5}$$

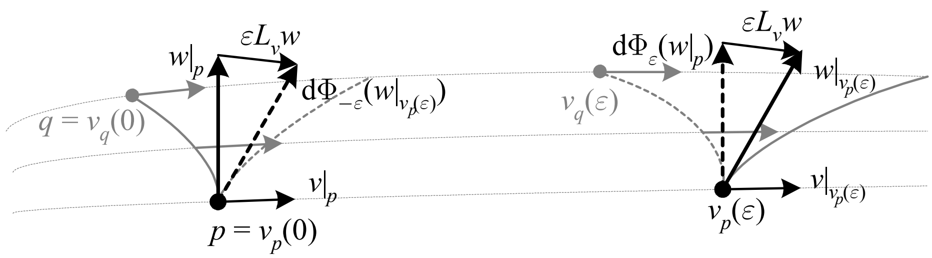

FIGURE C.2: The Lie derivative $L_v w$ is "the difference between w and its transport by the local flow of v."

☼ In this and future depictions of vector derivatives, the situation is simplified by focusing on the change in the vector field w while showing the "transport" of w as a parallel displacement. This has the advantage of highlighting the equivalency of defining the derivative at either 0 or ε in the limit $\varepsilon \to 0$. Depicting $L_v w$ as a non-parallel vector at $v_p(t)$ would be more accurate, but would obscure this fact. We also will follow the picture here in using words to characterize derivatives: namely, "the difference" is short for "the difference per unit ε to order ε in the limit $\varepsilon \to 0$."

This definition can be shown to be equivalent to $L_v w \equiv [v, w]$. Another way of depicting the Lie derivative that highlights the anti-commutativity of the Lie bracket is to consider $L_v w$ in terms of a loop defined by the flows of v and w.

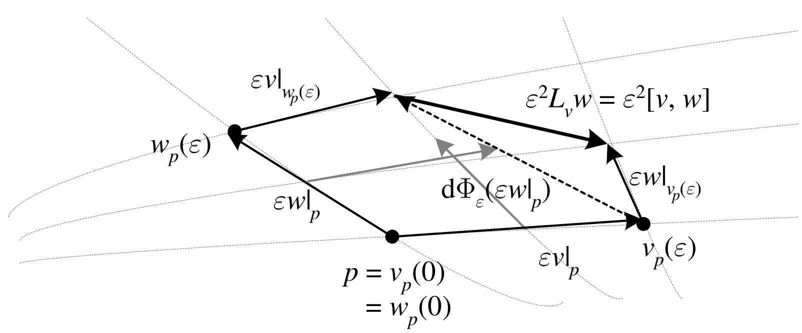

FIGURE C.3: The Lie derivative $L_v w$ can also be pictured as the vector field whose local flow is the "commutator of the flows of v and w," i.e. it is the difference between the local flow of v followed by w and that of w followed by v. Thus $L_v w$ "completes the parallelogram" formed by the flow lines.

C.3 The Lie derivative of an exterior form

The Lie derivative L_v can be applied to a k-form φ by using the pullback of φ by the diffeomorphism Φ associated with the flow of v, i.e. applied to k vectors $w_1, \ldots, w_k$ we define

$$L_v \varphi \left(w_1, \ldots, w_k\right) \equiv \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\varphi \left(\mathrm{d}\Phi_\varepsilon \left(w_1, \ldots, w_k\right)\right) - \varphi \left(w_1, \ldots, w_k\right)\right]. \tag{C.6}$$

$L_v \varphi$ thus measures the change in φ as its arguments are transported by the local flow of v. In the case of a 0-form f, this is just the differential or directional derivative $L_v f = v(f) = \mathrm{d}f(v)$.

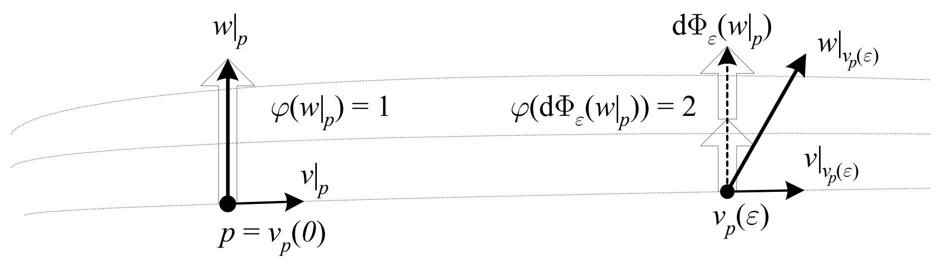

FIGURE C.4: The Lie derivative illustrated for a 1-form φ with $\varepsilon = 1$. $L_v\varphi$ is "the difference between φ applied to w and φ applied to w transported by the local flow of v," so above we have $L_v\varphi(w) = 2 - 1 = 1$ (valid in the limit $\varepsilon \to 0$ if φ changes linearly in the range shown).

☼ Here and in future figures, we represent a 1-form φ as a "receptacle" $\varphi^{\Uparrow} \equiv \varphi^{\sharp} / \left\|\varphi^{\sharp}\right\|^2$ which when applied to a vector "arrow" argument v yields the number of receptacles covered by the projection of v onto $\varphi^{\sharp}$, which is the value of $\varphi(v)$. This can be seen by recalling from Section A.1 that $\varphi(v) / \left\|\varphi^{\sharp}\right\|$ is the length of the projection of v onto $\varphi^{\sharp}$, so that this projection divided by the length of the receptacle $\left\|\varphi^{\Uparrow}\right\| = 1 / \left\|\varphi^{\sharp}\right\|$ recovers the value $\varphi(v)$. The advantage of this approach is that values can be calculated from the figure absent a length scale. Another common graphical device is to represent 1-forms as "surfaces" which are "pierced" by the arrows.

△ The common practice of depicting a 1-form φ in terms of the associated vector $\varphi^{\Uparrow}$ as above has consequences that can be non-intuitive. For example, doubling the value of the 1-form means halving its length in the illustration, i.e. the value of the 1-form can be viewed as the "density" of receptacles. Thus, when depicting φ as changing linearly, the length L of the 1-form representation changes like $L \mapsto L/(1 + r\varepsilon)$ for some scaling factor r, which doesn't appear linear as a vector representation would, whose length changes like $L \mapsto L(1 + r\varepsilon)$.

By using the above definitions of the Lie derivative applied to vectors and 1-forms, and noting that we can derive a Leibniz rule over contraction $L_v(\varphi\left(w\right)) = \left(L_v\varphi\right)\left(w\right) + \varphi\left(L_v w\right)$, we arrive at an expression for the Lie derivative applied to general tensors, viewed as real-valued mappings on vectors and 1-forms:

$$\begin{aligned}L_vT(\varphi_1,\dots,\varphi_m,w_1,\dots,w_n) &= v\left(T(\varphi_1,\dots,\varphi_m,w_1,\dots,w_n)\right)\\&-\sum_{j=1}^{m}T\left(\varphi_1,\dots,L_v\varphi_j,\dots,\varphi_m,w_1,\dots,w_n\right)\\&-\sum_{j=1}^{n}T\left(\varphi_1,\dots,\varphi_m,w_1,\dots,L_vw_j,\dots,w_n\right)\end{aligned} \tag{C.7}$$

In a holonomic frame, this yields an expression for the Lie derivative of a tensor in terms of coordinates

$$\begin{aligned}L_vT^{\mu_1\dots\mu_m}{}_{\sigma_1\dots\sigma_n} &= v^\lambda\frac{\partial}{\partial x^\lambda}T^{\mu_1\dots\mu_m}{}_{\sigma_1\dots\sigma_n}\\&-\sum_{j=1}^{m}\left(\frac{\partial v^{\mu_j}}{\partial x^\lambda}\right)T^{\mu_1\dots\mu_{j-1}\lambda\mu_{j+1}\dots\mu_m}{}_{\sigma_1\dots\sigma_n}\\&+\sum_{j=1}^{n}\left(\frac{\partial v^\lambda}{\partial x^{\sigma_j}}\right)T^{\mu_1\dots\mu_m}{}_{\sigma_1\dots\sigma_{j-1}\lambda\sigma_{j+1}\dots\sigma_n}.\end{aligned} \tag{C.8}$$

From this we can confirm that the Lie derivative satisfies the Leibniz rule over the tensor product, and therefore is a derivation of degree 0 on both the tensor algebra and the exterior algebra.

C.4 The exterior derivative of a 1-form

The Lie derivative $L_v\varphi$ is defined in terms of a vector field v, and its value as a "change in φ" is computed by using v to transport the arguments of φ. In contrast, recall that the differential d takes a 0-form $f\colon M\to\mathbb{R}$ to a 1-form $\mathrm{d}f\colon TM\to\mathbb{R}$ with

$$\mathrm{d}f(v)=v(f). \tag{C.9}$$

Thus d is a derivation of degree +1 on 0-forms, whose value as a "change in f" is computed using the vector field argument of the resulting 1-form.

We would like to generalize d to k-forms by extending this idea of including the "direction argument" by increasing the degree of the form. It turns out that if we also require the property

$$\mathrm{d}\left(\mathrm{d}\left(\varphi\right)\right)=0 \tag{C.10}$$

(or "$\mathrm{d}^2=0$"), there is a unique graded derivation of degree +1 that extends d to general k-forms; this derivation is called the **exterior derivative**. We first explore the exterior derivative of a 1-form.

The exterior derivative of a 1-form is defined by

$$\mathrm{d}\varphi\left(v,w\right)\equiv v\left(\varphi\left(w\right)\right)-w\left(\varphi\left(v\right)\right)-\varphi\left([v,w]\right), \tag{C.11}$$

where e.g.

$$v\left(\varphi\left(w\right)\right)=\lim_{\varepsilon\to0}\frac{1}{\varepsilon}\left[\varphi\left(w\left|_{v_p(\varepsilon)}\right.\right)-\varphi\left(w\left|_p\right.\right)\right] \tag{C.12}$$

measures the change in $\varphi\left(w\right)$ in the direction v, so that

$$
\begin{aligned}
\mathrm{d}\varphi\left(v, w\right) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \Big[& \left(\varphi\left(\varepsilon w\left|_{v_p(\varepsilon)}\right.\right) - \varphi\left(\varepsilon w\left|_p\right.\right)\right) \\
& - \left(\varphi\left(\varepsilon v\left|_{w_p(\varepsilon)}\right.\right) - \varphi\left(\varepsilon v\left|_p\right.\right)\right) \\
& - \varphi\left(\varepsilon^2\left[v, w\right]\right)\Big] .
\end{aligned}
\tag{C.13}
$$

The term involving the Lie bracket "completes the parallelogram" formed by v and w, so that $\mathrm{d}\varphi\left(v, w\right)$ can be viewed as the "sum of φ on the boundary of the surface defined by its arguments."

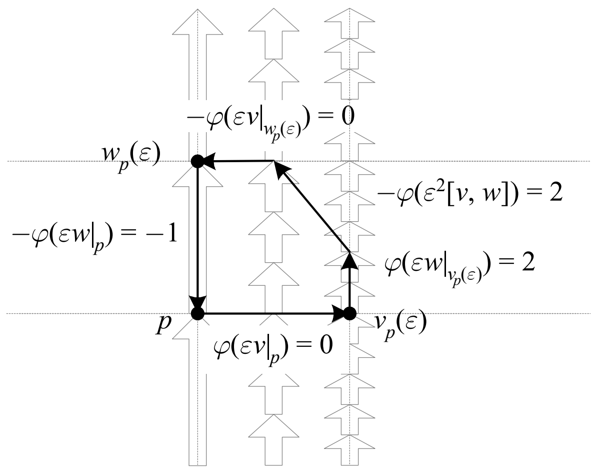

FIGURE C.5: The exterior derivative of a 1-form $\mathrm{d}\varphi\left(v, w\right)$ is the sum of φ along the boundary of the completed parallelogram defined by v and w. So if in the diagram $\varepsilon = 1$, we have $\mathrm{d}\varphi\left(v, w\right) = (2 - 1) - (0 - 0) + 2 = 3$. This value is valid in the limit $\varepsilon \to 0$ if the sum varies like ε^2 as depicted in the figure.

The identity $\mathrm{d}^2 = 0$ can then be seen as stating the intuitive fact that the boundary of a boundary is zero. If $\varphi = \mathrm{d}f$, then $\varphi\left(v\right) = \mathrm{d}f\left(v\right) = v\left(f\right)$, the change in f along v. Thus e.g. $\varepsilon\varphi\left(v\left|_p\right.\right) = f\left(v_p\left(\varepsilon\right)\right) - f\left(p\right)$, so that the value of φ on v is the difference between the values of f on the two points which are the boundary of v. Each endpoint will be cancelled by a starting point as we add up values of φ along a sequence of vectors, resulting in the difference between the values of f at the boundary of the total path defined by these vectors. $\mathrm{d}\varphi$ is the value of φ over the boundary path of the surface defined by its arguments, which has no boundary points and so vanishes.

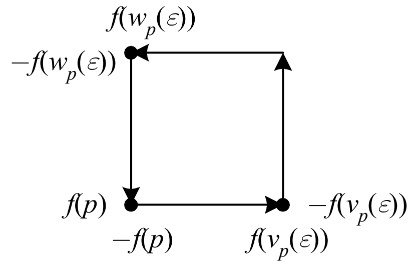

FIGURE C.6: $\mathrm{d}^2 = 0$ corresponds to the boundary of a boundary is zero: each term $\varphi(v) = \mathrm{d}f(v)$ is the difference between the values of f on the boundary points of v, which cancel as we traverse the boundary of the surface defined by the arguments of $\mathrm{d}\varphi(v, w)$. In the figure we assume a vanishing Lie bracket for simplicity.

Note that $\mathrm{d}\varphi\,(v, w)$ measures the interaction between φ and the vector fields v and w, thus avoiding the need to "transport" vectors. In particular, a non-zero exterior derivative can be pictured as resulting from either the vector fields or φ "changing," i.e. changing with regard to the implied coordinates of our pictures.

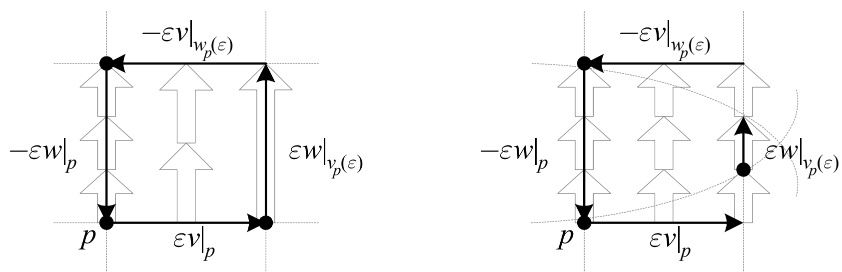

FIGURE C.7: A non-zero exterior derivative $\mathrm{d}\varphi\,(v, w)$ results from changes in $\varphi\,(v)$ or $\varphi\,(w)$, not changes in either φ or the vector fields alone as compared to some transport.

If we calculate $\mathrm{d}\varphi\,(e_1, e_2)$ explicitly in a holonomic frame in two dimensions, $\mathrm{d}\left(\varphi_1 \mathrm{d}x^1 + \varphi_2 \mathrm{d}x^2\right) = \mathrm{d}\varphi_1 \wedge \mathrm{d}x^1 + \mathrm{d}\varphi_2 \wedge \mathrm{d}x^2$, so applying this to the basis vector fields e_1 and e_2 we have

$$\begin{aligned}
\mathrm{d}\varphi\,(e_1, e_2) &= \mathrm{d}\varphi_1\,(e_1) \cdot \mathrm{d}x^1\,(e_2) - \mathrm{d}\varphi_1\,(e_2) \cdot \mathrm{d}x^1\,(e_1) \\
&\quad + \mathrm{d}\varphi_2\,(e_1) \cdot \mathrm{d}x^2\,(e_2) - \mathrm{d}\varphi_2\,(e_2) \cdot \mathrm{d}x^2\,(e_1) \\
&= e_1\,(\varphi_2) - e_2\,(\varphi_1) \\
&= \partial/\partial x^1\,(\varphi_2) - \partial/\partial x^2\,(\varphi_1)\,.
\end{aligned} \tag{C.14}$$

Note that a holonomic dual frame $\beta^\mu = \mathrm{d}x^\mu$ satisfies $\mathrm{d}\beta^\mu = \mathrm{dd}x^\mu = 0$.

C.5 The exterior derivative of a k-form

The extension of the coordinate-free definition of d to general k-forms gives the expression

$$\begin{aligned} &\mathrm{d}\varphi\left(v_0,\ldots,v_k\right) \\ &\equiv \sum_{j=0}^{k}(-1)^j\, v_j\left(\varphi\left(v_0,\ldots,v_{j-1},v_{j+1},\ldots,v_k\right)\right) \\ &\quad + \sum_{i<j}(-1)^{i+j}\,\varphi\left(\left[v_i,v_j\right],v_0,\ldots,v_{i-1},v_{i+1},\ldots,v_{j-1},v_{j+1},\ldots,v_k\right). \end{aligned} \tag{C.15}$$

Our picture of $\mathrm{d}^2=0$ for 1-forms then can be extended to higher dimensions. For example, assuming vanishing Lie brackets to simplify the picture, the exterior derivative of a 2-form $\mathrm{d}\varphi\left(u,v,w\right)$ can be viewed as the "sum of φ on the boundary faces of the cube defined by its arguments." If $\varphi=\mathrm{d}\psi\left(v,w\right)$ is the boundary of a face, $\mathrm{d}\varphi=\mathrm{d}^2\psi$ is the sum of the boundaries of the faces; each edge is then counted by two faces with opposite signs, thus canceling and confirming that $\mathrm{d}^2=0$.

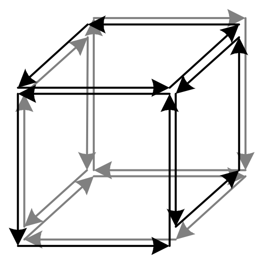

FIGURE C.8: The 3-form $\mathrm{d}\varphi=\mathrm{d}^2\psi$ sums ψ over the edges of the faces of a cube. The sum vanishes since each edge is counted twice with opposite signs.

In a holonomic frame, we can obtain an expression for $\mathrm{d}\varphi$ in terms of coordinates

$$\mathrm{d}\varphi=\sum_{\mu_0<\cdots<\mu_k}\left(\sum_{j=0}^{k}(-1)^j\,\frac{\partial}{\partial x^{\mu_j}}\varphi_{\mu_0\ldots\mu_{j-1}\mu_{j+1}\ldots\mu_k}\right)\mathrm{d}x^{\mu_0}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}, \tag{C.16}$$

or even more explicitly,

$$\begin{aligned} \mathrm{d}\varphi=\sum_{\mu_0<\cdots<\mu_k}&\left(\frac{\partial}{\partial x^{\mu_0}}\varphi_{\mu_1\ldots\mu_k}-\frac{\partial}{\partial x^{\mu_1}}\varphi_{\mu_0\mu_2\ldots\mu_k}+\cdots\right. \\ &\left.+(-1)^k\,\frac{\partial}{\partial x^{\mu_k}}\varphi_{\mu_0\ldots\mu_{k-1}}\right)\mathrm{d}x^{\mu_0}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}. \end{aligned} \tag{C.17}$$

It is not hard to see that the exterior derivative commutes with the pullback, i.e. $\Phi^*\mathrm{d}\varphi=\mathrm{d}\Phi^*\varphi$.

△ Despite a convenient description using coordinates associated with a holonomic frame, it is important to keep in mind that the exterior derivative of a form is frame- and coordinate-independent.

If we include an inner product, vector calculus can be seen to correspond to exterior calculus on $\mathbb{R}^3$, and can thus be generalized to arbitrary dimensions:

- For a function (0-form) f, the components of the 1-form $\mathrm{d}f$ correspond to those of the gradient of f, i.e. $(\mathrm{d}f)_\mu = (\nabla f)^\mu$ or $\nabla f = (\mathrm{d}f)^\sharp$; a generalization of the gradient is then the 1-form $\mathrm{d}f$

- For a 1-form with components equal to those of a vector $\varphi_\mu = v^\mu$, the components of $\mathrm{d}\varphi$ correspond to those of the curl of v, i.e. $(\mathrm{d}\varphi)_\mu = (\nabla \times v)^\mu$ or $(\nabla \times v) = (*\mathrm{d}(v^\flat))^\sharp$; a generalization of the curl is then the 2-form $\mathrm{d}\varphi$

- For a 2-form with components equal to those of a vector $\psi_\mu = (*\varphi)_\mu = v^\mu$, the value of $\mathrm{d}\psi$ corresponds to the value of the divergence of v, i.e. $\mathrm{d}\psi = \nabla \cdot v$ or $\nabla \cdot v = *\mathrm{d}(*(v^\flat))$; a generalization of the divergence is then the value $*\mathrm{d}(*\varphi)$

In $\mathbb{R}^3$ the relations curl grad = div curl = 0 thus correspond to the property $\mathrm{d}^2 = 0$. Note that we have used the musical isomorphisms on $\mathbb{R}^3$, which imply an inner product, as does the Hodge star. The generalizations can be extended to a pseudo inner product with signature (r, s) by defining the divergence as $(-1)^s * \mathrm{d}(*v^\flat)$, which is then independent of both signature and orientation.

Finally, the classical gradient, curl, and divergence integral theorems in vector calculus are generalized to **Stokes' theorem**: for an $(n-1)$-form φ on a compact oriented manifold M^n with boundary ∂M,

$$\int_M \mathrm{d}\varphi = \int_{\partial M} \varphi. \tag{C.18}$$

This is essentially the integral form of the property $\mathrm{d}^2 = 0$: summing $\mathrm{d}\varphi$ over M can be pictured as summing φ over the boundaries of infinitesimal volumes, so that all internal boundaries cancel and what is left is φ over the outer boundary ∂M.

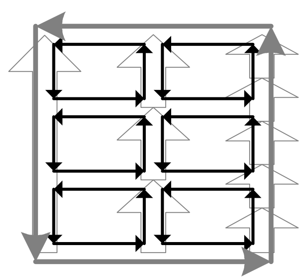

FIGURE C.9: The integral of $\mathrm{d}\varphi$ over M can be pictured as summing φ over the boundaries of infinitesimal volumes, so that all internal boundaries cancel and what is left is φ over the outer boundary ∂M.

C.6 Relationships between derivations

We can define one other derivation on k-forms, the **interior derivative** (AKA inner derivative, inner multiplication), which is the generalization of the interior product to forms on manifolds, i.e. for a given vector v it is the graded degree -1 derivation

$$(i_v\varphi)(w_2, \ldots, w_k) \equiv \varphi(v, w_2, \ldots, w_k) \tag{C.19}$$

on k-forms φ, which follows the graded Leibniz rule

$$i_v(\varphi \wedge \psi) = (i_v\varphi) \wedge \psi + (-1)^k \varphi \wedge (i_v\psi). \tag{C.20}$$

The graded commutativity of forms immediately gives the property $i_v i_w + i_w i_v = i_v^2 = 0$. We define $i_v f \equiv 0$ for a 0-form f and note that $i_v\Omega = *(v^\flat)$.

The interior, exterior, and Lie derivatives then form an infinite-dimensional graded Lie algebra with the following relations:

- $[L_v, L_w] \equiv L_v L_w - L_w L_v = L_{[v,w]}$
- $[i_v, i_w] \equiv i_v i_w + i_w i_v = 0$
- $[\mathrm{d}, \mathrm{d}] \equiv \mathrm{d}^2 + \mathrm{d}^2 = 0$
- $[L_v, i_w] \equiv L_v i_w - i_w L_v = i_{[v,w]}$
- $[L_v, \mathrm{d}] \equiv L_v \mathrm{d} - \mathrm{d} L_v = 0$
- $[i_v, \mathrm{d}] \equiv i_v \mathrm{d} + \mathrm{d} i_v = L_v$

This last relation is sometimes called **Cartan's formula** (AKA Cartan's magic formula).

References

[1] M. do Carmo, *Riemannian Geometry* (Birkhï¿½user Boston, 1992)

[2] T. Frankel, *The Geometry of Physics* (Cambridge University Press, 1997)

[3] T. Frankel, *Gravitational Curvature* (W H Freeman and Company, 1979)

[4] M. Gï¿½ckeler and T. Schï¿½cker, *Differential Geometry, Gauge Theories, and Gravity* (Cambridge University Press, 1987)

[5] A. Gray, Alfred, "The volume of a small geodesic ball of a Riemannian manifold," The Michigan Mathematical Journal 20 (1974), no. 4, 329-344, http://projecteuclid.org/euclid.mmj/1029001150

[6] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (John Wiley & Sons, 1963)

[7] C. Misner, K. Thorne and J. Wheeler, *Gravitation* (W H Freeman and Company, 1973)

[8] P. Petersen, *Riemannian Geometry* (Springer, 2006)

[9] R. Wald, *General Relativity* (University of Chicago Press, 1984)