Connections on fibre bundles

Virinchi Rallabhandi

April 2022

Contents

Whether we recognise them as such or not, connections on fibre bundles appear quite frequently in modern theoretical physics. This treatise is intended to be the document I wish I had as a theoretical physics student - and thus takes a unique route through the subject. Rather than present Riemannian geometry, Yang-Mills theory and spin connections as philosophically similar but seemingly distinct subjects, I have endeavoured to present them all under the banner of a single unifying theory. As usual, abstraction is the price of generality, but I think in this case it is well worth paying. Unfortunately, unless I eventually expand these musings into a complete book, the reader may require a somewhat eclectic base of pre-requisite differential geometry knowledge to comprehend what follows. As such this presentation is not very useful without a prior "first-pass" exposure to some of the topics I discuss. Nothing I present here is original - my only contribution is to collate results spread across various sources and various sections within each source. I have relied heavily on Jack Smith's Cambridge Part III course on differential geometry and Mikio Nakahara's "Geometry, topology and physics."

1 General theory

Before I plough ahead with connections and fibre bundles, I'll clarify what I mean by a vector in a tangent space because my favoured definition is slightly non-standard (albeit ultimately equivalent to the standard one).

Definition 1.1 (Curve based at p). *A curve based at $p \in M$ is defined to be a smooth map, $\gamma : I \to M$, where I is a connected open neighbourhood of $\mathbb{R}$ and $\gamma(0) = p$.*

Definition 1.2 (Agreement to 1st order). *Two curves based at p, γ_1 and γ_2, are defined to agree to 1st order at p if and only if $\exists$ a chart, say $\varphi : V \to W$, about p such that*

$$\frac{\mathrm{d}\varphi(\gamma_1(t))}{\mathrm{d}t}\Big|_{t=0} = \frac{\mathrm{d}\varphi(\gamma_2(t))}{\mathrm{d}t}\Big|_{t=0} \tag{1}$$

as vectors in $\mathbb{R}^{\dim(M)}$.

Corollary 1.2.1. *If the condtion above holds for some chart about p, then it holds for all charts about p.*

Proof. Let $\dim(M) = n$. For a chart, $\varphi : V \to W$, about p, let $\pi_p^\varphi : \{\text{curves based at } p\} \to \mathbb{R}^n$ be defined by

$$\pi_p^\varphi : \gamma \mapsto \frac{\mathrm{d}\varphi(\gamma(t))}{\mathrm{d}t}\bigg|_{t=0} = (\varphi \circ \gamma)'(0) \tag{2}$$

Let φ_1 and φ_2 be two different charts about p.

$$\begin{aligned}
\therefore \pi_p^{\varphi_2}(\gamma) &= (\varphi_2 \circ \gamma)'(0) && (3)\\
&= (\varphi_2 \circ \varphi_1^{-1} \circ \varphi_1 \circ \gamma)'(0) && (4)\\
&= (\varphi_2 \circ \varphi_1^{-1})'((\varphi_1 \circ \gamma)(0))(\varphi_1 \circ \gamma)'(0) \text{ by the chain rule of } \mathbb{R}^n && (5)\\
&= A \circ \pi_p^{\varphi_1}(\gamma) && (6)
\end{aligned}$$

where $A = (\varphi_2 \circ \varphi_1^{-1})'(\varphi_1(p))$. Since A is invertible (the inverse is $(\varphi_1 \circ \varphi_2^{-1})'(\varphi_2(p)))$, it follows that $\pi_p^{\varphi_2}(\gamma_1) = \pi_p^{\varphi_2}(\gamma_2) \iff \pi_p^{\varphi_1}(\gamma_1) = \pi_p^{\varphi_1}(\gamma_2)$. □

Corollary 1.2.2. *Agreement to 1st order is an equivalence relation on the set of curves based at p.*

Definition 1.3 (Tangent space)**.** *The tangent space to M at p, denoted T_pM, is defined to be $\{$curves based at $p\}$/agreement to first order, i.e. $T_pM = \{[\gamma]\}$. The equivalence classes, $[\gamma]$, are called vectors (at p).*

This definition can be shown to be completely equivalent to the standard definition using derivations at p, although I think it would be too much of a digression to prove that here. However, the definition presented here is better for a number of reasons. Philosophically, it is much more geometric in nature - I think that better aligns with the spirit of differential *geometry* than introducing derivations at each point, which I find unnecessarily abstract. At a more low-brow level, the equivalence class definition makes it manifest that $\dim(T_pM) = \dim(M)$, where as one would intuitively think that the space of derivations is horribly infinite dimensional (because the manifolds are only assumed to be smooth, not analytic). From the equivalence class point of view, the coordinate basis, $\{\partial_i\}_{i=1}^{\dim(M)}$, in some chart, φ, would be $\{(\pi_p^\varphi)^{-1}(e_i)\}_{i=1}^{\dim(M)}$ where $\{e_i\}_{i=1}^{\dim(M)}$ is the standard basis of $\mathbb{R}^{\dim(M)}$ and π_p^φ is defined by equation 2. Finally, the equivalence class definition also makes the pushforward much more transparent.

Definition 1.4 (Pushforward/tangent map/dragging/differential/derivative/total derivative/insertYourOwnNomenclatureHere)**.** *Let $f : M \to N$ be a smooth map. Then, the pushforward of f at $p \in M$ is defined to be the linear map, $f_* : T_pM \to T_{f(p)}N$, given by $f_*[\gamma] = [f \circ \gamma]$.*

I will also deploy the following convention for defining tangent vectors to curves. If a vector, v, is given by $[\gamma]$, then I will also write $v = \frac{\mathrm{d}\gamma(t)}{\mathrm{d}t}|_{t=0}$. For example, the integral curves, $\gamma(t)$, of a vector field, $v(p)$, would satisfy $\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t} = v(\gamma(t))$.

Having addressed these administrative wrinkles and assumed a sufficient background in differential geometry, I am ready to zoom ahead to the main topics I wish to present.

Definition 1.5 (Differentiable fibre bundle)**.** *A differentiable fibre bundle is a collection, (E, B, F, π, G), such that*

1. *E is a manifold, called the total space.*
2. *B is a manifold, called the base.*
3. *F is a manifold, called the fibre.*
4. *$\pi : E \to B$ is a smooth surjection, called the projection.*
5. *G is a Lie group, called the structure group, having some specific left group action on F.*
6. *$\exists$ an open cover, $\{V_\alpha\}_{\alpha \in A}$, of B such that $\forall \alpha$, $\exists$ a diffeomorphism, $\Phi_\alpha : \pi^{-1}(V_\alpha) \to V_\alpha \times F$. The V_α and Φ_α are together called a local trivialisation.*
7. *$\forall \alpha$, $\mathrm{pr}_1 \circ \Phi_\alpha = \pi$, where pr_1 is projection on the first set in a cartesian product.*
8. *$\forall \alpha, \beta$, the map, $\Phi_\beta \circ \Phi_\alpha^{-1} : (V_\alpha \cap V_\beta) \times F \to (V_\alpha \cap V_\beta) \times F$, is of the form, $(p, f) \mapsto (p, g_{\beta\alpha}(p) \cdot f)$, for some smooth functions, $g_{\beta\alpha} : V_\alpha \cap V_\beta \to G$, called the transition functions.*

In practice, it is impossible to present this wealth of data in any compact notation. Thus, when the contexts are clear, a differentiable fibre bundle will simply be presented as $\pi : E \to B$.

Definition 1.6 (Section)**.** *Given a differential fibre bundle, $\pi : E \to B$, a section is defined to be a smooth map, $s : B \to E$, such that $\pi \circ s = \mathrm{id}_B$. A local section is defined in exactly the same way, except that the domain of s only has to be some open set, $V \subseteq B$.*

Sections will be indispensable later. But now, I would like to introduce two particular types of differentiable fibre bundle that will play a starring role in the theory of connections.

Definition 1.7 (Principal G-bundle)**.** *A principal G-bundle is a differential fibre bundle with structure group and fibre both G. Furthermore, the left group action must be left (group) multiplication.*

Henceforth, I will abbreviate "principal G-bundle" to just "principal bundle" or "G-bundle" depending on context and the whims of my taste. Every G-bundle carries a canonical right G-action.

Definition 1.8 (Canonical right G-action)**.** *Given a G-bundle, $\pi : P \to B$, the canonical right G-action is defined in trivialisations as follows.*
Let $p \in P$, let Φ_α be a local trivialisation around p and let $\Phi_\alpha(p) = (b, g)$. Then, $h \in G$ is defined to act by $p \cdot h = \Phi_\alpha^{-1}(b, gh)$.

Corollary 1.8.1. *The canonical right G-action is independent of the the choice of trivialisation because the transition functions act on the left.*

Corollary 1.8.2. *The canonical right G-action is free.*

There is also an infinitesimal version of this group action.

Definition 1.9 (Canonical right $\mathfrak{g}$-action)**.** *Given a G-bundle, $\pi : P \to B$, $\xi \in \mathfrak{g}$ is defined to act by $p \cdot \xi = \frac{\mathrm{d}}{\mathrm{d}t}(p \cdot \mathrm{e}^{t\xi})|_{t=0}$, where $\mathrm{e}^{\cdot}$ is the exponential map. Thus, the output, $p \cdot \xi$, is an element of T_pP.*

Theorem 1.10. *For a principal bundle, $\pi : P \to B$, the canonical right G-action provides a correspondence between local trivialisations and local sections.*

Proof. Let Φ_α be a local trivialisation over a trivialising patch, V_α. Then, I can define a local section, $s_\alpha : V_\alpha \to P$, by $s_\alpha(b) = \Phi_\alpha^{-1}(b, e)$. This really is a local section because s_α is manifestly smooth and $(\pi \circ s_\alpha)(b) = (\pi \circ \Phi_\alpha^{-1})(b, e) = \mathrm{pr}_1(b, e) = b$.
Conversely, let $s : V \to P$ be a local section. Then, I can define $\phi : V \times G \to P$ by $\phi(b, g) = s(b) \cdot g$.
Then, $\phi(b, g) = \phi(b, h) \iff s(b) \cdot g = s(b) \cdot h \iff g = h$ because the group action is free.
$\therefore \phi$ is injective. It also manifestly has all possible smoothness properties.
$\therefore \{s(b) \cdot g \,|\, b \in V, g \in G\}$ is in bijection with $\{(b, g) \,|\, b \in V, g \in G\} = V \times G$.
As P is a principal bundle, there must exist some trivialisation, φ, around $\pi^{-1}(b)$. In this trivialisation, $s(b) \cdot g = \varphi^{-1}(b, g)$. But, because φ is a diffeomorphism on some open neighbourhood of $\pi^{-1}(b)$, varying g means $s(b) \cdot g$ varies over all of $\pi^{-1}(b)$.
$\therefore$ The ϕ above provides a diffeomorphism between $\pi^{-1}(b)$ and $V \times G$.
$\therefore \phi^{-1}$ is the required local trivialisation. □

Corollary 1.10.1. *A principal bundle is trivial (i.e. diffeomorphic to $B \times G$) if and only if it admits a global section.*

The other prominent example of a differentiable fibre bundle is the vector bundle.

Definition 1.11 (Vector bundle)**.** *A rank-k vector bundle is a differential fibre bundle where the fibre is $\mathbb{R}^k$, the structure group is $\mathrm{GL}(k, \mathbb{R})$ and the left group action is matrix multiplication.*

Although I have introduced them very generally, examples of vector bundles are almost certainly already familiar to any reader who satisfies the pre-requisites I'm assuming. In particular, I'll assume (although it won't be needed unti section 2) the reader is familiar with tangent bundles, cotangent bundles and the various tensor bundles generated from these. Furthermore, it can be shown using the "pseudo-atlas" construction that specifying the transition functions (and thus trivialising neighbourhoods) is sufficient for uniquely determining a vector or principal bundle (up to an appropriate notion of isomorphism). This is a point I do not wish to belabour here.

The central concept underpinning this treatise is connections on principal bundles.

Definition 1.12 (Connection on a principal bundle)**.** *Let $\pi : P \to B$ be a G-bundle. A connection on P is defined to be a $\mathfrak{g}$-valued 1-form, $\mathcal{A}$, on P (i.e. $\mathcal{A} \in \mathfrak{g} \otimes \Gamma(T^*P)$ where Γ denotes the set of sections) such that*

1. $\mathcal{A}_p(p \cdot \xi) = \xi \;\; \forall p \in P,\, \xi \in \mathfrak{g}$.
2. $(R_g)^*\mathcal{A} = \mathrm{Ad}_{g^{-1}}\mathcal{A} \;\; \forall g \in G$.

This definition requires a fair bit of interpretation (so here it is ...).
$\mathcal{A}_p$ denotes $\mathcal{A}$ at the point, p, and it's the covector part of $\mathcal{A}_p$ that acts on $p \cdot \xi$, with the "$\mathfrak{g}$-valued" part just coming along for the ride.
$(R_g)^*$ is the pullback of $\mathcal{A}$ under the canonical right G-action, i.e. $R_g : P \to P$ by $R_g : p \mapsto p \cdot g$.
Again, the "$\mathfrak{g}$-valued" part of $\mathcal{A}$ just comes along for the ride.
In my conventions, the adjoint representation is defined to be the pushforward,
$\mathrm{Ad}_g(\xi) = (C_g)_*\xi$, where $g \in G$, $\xi \in \mathfrak{g}$ and for any $h \in G$, $C_g(h) = ghg^{-1}$. Thus, in $\mathrm{Ad}_{g^{-1}}\mathcal{A}$, the $\mathrm{Ad}_{g^{-1}}$ acts on the "$\mathfrak{g}$-valued" part and the 1-form part comes along for the ride.

At this point, a connection looks ridiculously esoteric and unuseful. But, it's power is slowly revealed with the correct/lucky exploration. In the physics community, connections are typically defined using their local versions, and corresponding properties, on the base space.

Definition 1.13 (Local connection 1-form). *Let Φ_α be a local trivialisation of P corresponding to a local section, s_α. Then, the local connection 1-form is defined to be $A_\alpha = (s_\alpha)^*\mathcal{A}$. Thus, $A_\alpha \in \mathfrak{g} \otimes \Gamma(T^*V_\alpha)$.*

Theorem 1.14. *On overlaps of trivialising patches, the local connection 1-forms are related by*[1] $A_\alpha = g_{\beta\alpha}^{-1}\mathrm{d}g_{\beta\alpha} + \mathrm{Ad}_{g_{\beta\alpha}^{-1}}A_\beta \iff A_\beta = g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} + \mathrm{Ad}_{g_{\beta\alpha}}A_\alpha$.

Proof. Before boldly going into the proof, it's best elucidate the theorem's meaning.
$\mathrm{d}g_{\beta\alpha}$ is a local 1-form with values in $T_{g_{\beta\alpha}}G$, i.e. $\mathrm{d}g_{\beta\alpha} \in T_{g_{\beta\alpha}}G \otimes \Gamma(T^*V_\alpha)$.
$g_{\beta\alpha}^{-1}\mathrm{d}g_{\beta\alpha}$ is shorthand[2] for $(L_{g_{\beta\alpha}^{-1}})_*\mathrm{d}g_{\beta\alpha}$.
$\therefore$ The pushforward gives an output in $T_eG \otimes \Gamma(T^*V_\alpha) = \mathfrak{g} \otimes \Gamma(T^*V_\alpha)$, as required.
Let Φ_α be the trivialisation corresponding to s_α. Then, for any $b \in V_\alpha \cap V_\beta$ and $h \in G$, $(\Phi_\beta \circ \Phi_\alpha^{-1})(b,h) = (b, g_{\beta\alpha}(b)h)$.
However, $(\Phi_\beta \circ \Phi_\alpha^{-1})(b,h) = \Phi_\beta(s_\alpha(b)\cdot h)$
$\therefore \Phi_\beta(s_\alpha(b)\cdot h) = (b, g_{\beta\alpha}(b)h) = \Phi_\beta(s_\beta(b)\cdot(g_{\beta\alpha}(b)h))$.
$\therefore s_\alpha(b)\cdot h = s_\beta(b)\cdot(g_{\beta\alpha}(b)h)$ as Φ_β is injective.
$\therefore s_\alpha(b) = s_\beta(b)\cdot g_{\beta\alpha}(b)$ by acting with h^{-1} on both sides.
Let $[\gamma(t)]$ be an arbitrary element of T_bB. Then, by the definition of pushforward,

$$(s_\beta)_*[\gamma] = (s_\alpha \cdot g_{\beta\alpha}^{-1})_*[\gamma] \tag{7}$$
$$= [(s_\alpha \cdot g_{\beta\alpha}^{-1}) \circ \gamma] \tag{8}$$
$$= [(s_\alpha \circ \gamma)\cdot g_{\beta\alpha}^{-1}(\gamma)] \tag{9}$$

The equivalence classes are defined by their derivative at $t = 0$. Using primes for derivatives,

$$((s_\alpha \circ \gamma)\cdot g_{\beta\alpha}^{-1}(\gamma))'(0) = (s_\alpha \circ \gamma)'(0)\cdot g_{\beta\alpha}^{-1}(\gamma(0)) + (s_\alpha \circ \gamma)(0)\cdot g_{\beta\alpha}^{-1}(\gamma)'(0) \tag{10}$$
$$= (s_\alpha \circ \gamma)'(0)\cdot g_{\beta\alpha}^{-1}(b) + s_\alpha(b)\cdot g_{\beta\alpha}^{-1}(\gamma)'(0) \tag{11}$$
$$= (s_\alpha \circ \gamma)'(0)\cdot g_{\beta\alpha}^{-1}(b) + s_\beta(b)\cdot g_{\beta\alpha}(b)\cdot g_{\beta\alpha}^{-1}(\gamma)'(0) \tag{12}$$
$$\therefore (s_\beta)_*[\gamma] = (R_{g_{\beta\alpha}^{-1}})_*[s_\alpha \circ \gamma] + (s_\beta(b)\cdot(g_{\beta\alpha}(b)\mathrm{d}g_{\beta\alpha}^{-1}))([\gamma]) \tag{13}$$
$$= (R_{g_{\beta\alpha}^{-1}})_*(s_\alpha)_*[\gamma] + (s_\beta(b)\cdot(g_{\beta\alpha}(b)\mathrm{d}g_{\beta\alpha}^{-1}))([\gamma]) \tag{14}$$
$$\therefore (s_\beta)_* = (R_{g_{\beta\alpha}^{-1}})_*(s_\alpha)_* + s_\beta(b)\cdot(g_{\beta\alpha}(b)\mathrm{d}g_{\beta\alpha}^{-1}) \text{ as } [\gamma] \text{ is arbitrary} \tag{15}$$

Finally, let v be an arbitrary element of T_bB (I won't need the equivalence class representation this time).

$$\therefore A_\beta(v) = ((s_\beta)^*\mathcal{A})(v) \tag{16}$$
$$= \mathcal{A}((s_\beta)_*v) \tag{17}$$
$$= \mathcal{A}((R_{g_{\beta\alpha}^{-1}})_*(s_\alpha)_*v + (s_\beta\cdot(g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}))v) \tag{18}$$
$$= (R_{g_{\beta\alpha}^{-1}})^*\mathcal{A}((s_\alpha)_*v) + \mathcal{A}((s_\beta\cdot(g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}))v) \tag{19}$$
$$= (\mathrm{Ad}_{g_{\beta\alpha}}\mathcal{A})((s_\alpha)_*v) + (g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1})(v) \text{ from the axioms of } \mathcal{A} \tag{20}$$
$$= (\mathrm{Ad}_{g_{\beta\alpha}}(s_\alpha)^*\mathcal{A})(v) + (g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1})(v) \tag{21}$$
$$= (\mathrm{Ad}_{g_{\beta\alpha}}A_\alpha)(v) + (g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1})(v) \tag{22}$$

$\therefore A_\beta = \mathrm{Ad}_{g_{\beta\alpha}}A_\alpha + g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}$, as v is arbitrary. □

[1] Some of the notation to follow is explained in the proof (meaning you have no choice but to read it).

[2] When G is a matrix Lie group, the pushforward is just matrix multiplication and so this is no longer shorthand, but literally what is intended.

This theorem encompasses the typical point of view on connections in the physics community. Luckily, it's also possible to go from here back to the mathematician's perspective.

Theorem 1.15. *Given a set of $\mathfrak{g}$-valued 1-forms, A_α, related to each other by*

$$A_\beta = \mathrm{Ad}_{g_{\beta\alpha}} A_\alpha + g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}, \tag{23}$$

$\exists$ a G-bundle, $\pi : P \to B$, such that the A_α are the local connection 1-forms of a connection, $\mathcal{A}$, on P.

Proof. The transition functions completely specify the bundle up to isomorphism, so a G-bundle with transition functions, $g_{\beta\alpha}$, exists.
I'll start by defining a connection, $\mathcal{A}_\alpha$, on $\pi^{-1}(V_\alpha)$.
Let $p \in \pi^{-1}(V_\alpha) \subseteq P$, let $b = \pi(p)$ and let $p = s_\alpha(b) \cdot g$. Then, define $\mathcal{A}_{\alpha,p}$ by

$$\mathcal{A}_{\alpha,p} = \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha + g^{-1}\mathrm{d}g \tag{24}$$

where d is the exterior derivative on P in this case.
Let v be an arbitrary element of $T_b V_\alpha$. Then, since $g = e$ for $p = s_\alpha(b)$,

$$\begin{aligned}
((s_\alpha)^*\mathcal{A}_{\alpha,s_\alpha(b)})(v) &= \mathcal{A}_{\alpha,s_\alpha(b)}((s_\alpha)_* v) && (25)\\
&= \pi^* A_\alpha((s_\alpha)_* v) + \mathrm{d}g((s_\alpha)_* v) && (26)\\
&= A_\alpha(\pi_*(s_\alpha)_* v) + \mathrm{d}g((s_\alpha)_* v) && (27)\\
&= A_\alpha(v) && (28)
\end{aligned}$$

since $\pi \circ s_\alpha = \mathrm{id} \implies \pi_*(s_\alpha)_* = \mathrm{id}$ and $\mathrm{d}g((s_\alpha)_* v) = 0$ as $g = e$ along $(s_\alpha)_* v$.
The upshot is that A_α really is the local connection 1-form of $\mathcal{A}_\alpha$.
Next, I have to show that on overlaps $\mathcal{A}_\alpha = \mathcal{A}_\beta$, thereby allowing me to define a $\mathcal{A}$ globally. From the transformation property of the A_α and $s_\alpha = s_\beta \cdot g_{\beta\alpha}$ (proved in the last theorem's proof),

$$\begin{aligned}
\mathcal{A}_\beta &= \mathrm{Ad}_{(g_{\beta\alpha}g)^{-1}}\pi^* A_\beta + (g_{\beta\alpha}g)^{-1}\mathrm{d}(g_{\beta\alpha}g) && (29)\\
&= \mathrm{Ad}_{(g_{\beta\alpha}g)^{-1}}\pi^*(\mathrm{Ad}_{g_{\beta\alpha}} A_\alpha + g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}) + g^{-1}g_{\beta\alpha}^{-1}\mathrm{d}(g_{\beta\alpha}g) && (30)
\end{aligned}$$

The π^* relates the 1-form parts on P and B, where as the Ad acts on the $\mathfrak{g}$-valued part.
$\therefore$ They commute.

$$\therefore \mathrm{Ad}_{(g_{\beta\alpha}g)^{-1}}\pi^*\mathrm{Ad}_{g_{\beta\alpha}} A_\alpha = \mathrm{Ad}_{g^{-1}g_{\beta\alpha}^{-1}}\pi^*\mathrm{Ad}_{g_{\beta\alpha}} A_\alpha = \mathrm{Ad}_{g^{-1}g_{\beta\alpha}^{-1}}\mathrm{Ad}_{g_{\beta\alpha}}\pi^* A_\alpha = \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha \tag{31}$$

As for the other terms, bearing in mind my notation that $g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}$ really means $(L_{g_{\beta\alpha}})_*\mathrm{d}g_{\beta\alpha}^{-1}$,

$$\begin{aligned}
g^{-1}g_{\beta\alpha}^{-1}\mathrm{d}(g_{\beta\alpha}g) &= g^{-1}g_{\beta\alpha}^{-1}\mathrm{d}(g_{\beta\alpha})g + g^{-1}g_{\beta\alpha}^{-1}g_{\beta\alpha}\mathrm{d}g && (32)\\
&= \mathrm{Ad}_{g^{-1}}g_{\beta\alpha}^{-1}\mathrm{d}(g_{\beta\alpha}) + g^{-1}\mathrm{d}g && (33)\\
&= -\mathrm{Ad}_{g^{-1}}\mathrm{d}(g_{\beta\alpha}^{-1})g_{\beta\alpha} + g^{-1}\mathrm{d}g && (34)\\
&= -\mathrm{Ad}_{g^{-1}}(R_{g_{\beta\alpha}})_*\mathrm{d}(g_{\beta\alpha}^{-1}) + g^{-1}\mathrm{d}g && (35)
\end{aligned}$$

There should be a pullback here for this expression to really make sense, because $\mathrm{d}g$ is a ($T_g G$-valued) 1-form on P, where as $\mathrm{d}g_{\beta\alpha}^{-1}$ is a ($T_{g_{\beta\alpha}^{-1}}G$-valued) 1-form on B. So really, the correct expression is

$$g^{-1}g_{\beta\alpha}^{-1}\mathrm{d}(g_{\beta\alpha}g) = -\mathrm{Ad}_{g^{-1}}(R_{g_{\beta\alpha}})_*\pi^*\mathrm{d}(g_{\beta\alpha}^{-1}) + g^{-1}\mathrm{d}g \tag{36}$$

Putting these expressions back into equation 30, I get

$$
\begin{aligned}
\mathcal{A}_\beta &= \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha + \mathrm{Ad}_{g^{-1}g_{\beta\alpha}^{-1}}\pi^* g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} - \mathrm{Ad}_{g^{-1}}(R_{g_{\beta\alpha}})_*\pi^*\mathrm{d}(g_{\beta\alpha}^{-1}) + g^{-1}\mathrm{d}g && (37)\\
&= \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha + \mathrm{Ad}_{g^{-1}}(R_{g_{\beta\alpha}})_*\pi^*\mathrm{d}g_{\beta\alpha}^{-1} - \mathrm{Ad}_{g^{-1}}(R_{g_{\beta\alpha}})_*\pi^*\mathrm{d}(g_{\beta\alpha}^{-1}) + g^{-1}\mathrm{d}g && (38)\\
&= \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha + g^{-1}\mathrm{d}g && (39)\\
&= \mathcal{A}_\alpha && (40)
\end{aligned}
$$

Finally, I have to check that the $\mathcal{A}$ I have defined using 24 actually satisfies the axioms of a connection. For that, let $\xi \in \mathfrak{g}$. Then,

$$
\begin{aligned}
\mathcal{A}_p(p\cdot\xi) &= \mathrm{Ad}_{g^{-1}}\pi^* A_\alpha(p\cdot\xi) + g^{-1}\mathrm{d}g(p\cdot\xi) && (41)\\
&= \mathrm{Ad}_{g^{-1}} A_\alpha(\pi_*(p\cdot\xi)) + g^{-1}\mathrm{d}g(p\cdot\xi) && (42)
\end{aligned}
$$

The first term has $\pi_*(p\cdot\xi) = \pi_*\frac{\mathrm{d}}{\mathrm{d}t}(p\cdot \mathrm{e}^{t\xi})|_{t=0} = \pi_*[p\cdot\mathrm{e}^{t\xi}] = [\pi(p\cdot\mathrm{e}^{t\xi})] = [0]$ since the canonical right G-action only acts within a fibre and so doesn't change the output under π. That leaves

$$
\begin{aligned}
\mathcal{A}_p(p\cdot\xi) &= g^{-1}\mathrm{d}g(p\cdot\xi) && (43)\\
&= (L_{g^{-1}})_*\mathrm{d}g([p\cdot\mathrm{e}^{t\xi}]) && (44)\\
&= (L_{g^{-1}})_*\frac{\mathrm{d}}{\mathrm{d}t}\left(g(p\cdot\mathrm{e}^{t\xi})\right)\Big|_{t=0} && (45)
\end{aligned}
$$

where $g(p\cdot\mathrm{e}^{t\xi})$ means the group element required to generate $p\cdot\mathrm{e}^{t\xi}$ by acting with the canonical right action on $s_\alpha(b)$. But, by definition, $p\in\pi^{-1}(b)$ is generated by g[3]. Thus,

$$
\begin{aligned}
\mathcal{A}_p(p\cdot\xi) &= (L_{g^{-1}})_*\frac{\mathrm{d}}{\mathrm{d}t}\left(g\mathrm{e}^{t\xi}\right)\Big|_{t=0} && (46)\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(g^{-1}g\mathrm{e}^{t\xi}\right)\Big|_{t=0} && (47)\\
&= \xi && (48)
\end{aligned}
$$

For the remaining axiom, let v be an arbitrary element of T_pP. Then, $(R_h)^*\mathcal{A}(v) = \mathcal{A}((R_h)_*v)$. The pushforward means that $(R_h)_*v$ is a vector in the tangent space of the point, $s_\alpha(b)\cdot(gh)$, not just $s_\alpha(b)\cdot g$.

$$
\begin{aligned}
\therefore\ (R_h)^*\mathcal{A}(v) &= \mathrm{Ad}_{(gh)^{-1}}\pi^* A_\alpha((R_h)_*v) + (gh)^{-1}\mathrm{d}(gh)((R_h)_*v) && (49)\\
&= \mathrm{Ad}_{h^{-1}g^{-1}} A_\alpha(\pi_*(R_h)_*v) + h^{-1}g^{-1}\mathrm{d}(gh)((R_h)_*v) && (50)
\end{aligned}
$$

The right G-action only acts within a fibre $\implies \pi\circ R_h = \pi \implies \pi_*(R_h)_* = \pi_*$. In the second term, h is just a constant, so $\mathrm{d}(gh)((R_h)_*v) = (R_h)_*(\mathrm{d}g(v))$.

$$
\begin{aligned}
\therefore\ (R_h)^*\mathcal{A}(v) &= \mathrm{Ad}_{h^{-1}g^{-1}} A_\alpha(\pi_*v) + h^{-1}g^{-1}(R_h)_*\mathrm{d}g(v) && (51)\\
&= \mathrm{Ad}_{h^{-1}}\mathrm{Ad}_{g^{-1}}\pi^* A_\alpha(v) + \mathrm{Ad}_{h^{-1}}g^{-1}\mathrm{d}g(v) && (52)\\
&= \mathrm{Ad}_{h^{-1}}\mathcal{A}(v) && (53)
\end{aligned}
$$

Thus, $(R_h)^*\mathcal{A} = \mathrm{Ad}_{h^{-1}}\mathcal{A}$ as v was arbitrary. □

I think a brief sanity check is also in order before proceeding further.

[3]Unfortunately, the same symbol, g, effectively has two different meanings in these expressions. It may be better to use g and g_0 to clarify this, but that will lead to its own notational irritations.

Theorem 1.16. *Every G-bundle, $\pi : P \to B$, admits a connection.*

Proof. Cover P with trivialisations, Φ_α, over V_α. Next, choose all the A_α to be zero. These define connections, $\mathcal{A}_\alpha$, on $\pi^{-1}(V_\alpha)$, but not the whole space and without being pieced together properly on overlaps. To get around that, define $\mathcal{A}$ globally by

$$\mathcal{A}_p = \sum_\alpha \rho_\alpha(\pi(p))\mathcal{A}_{\alpha,p} \tag{54}$$

where $\{\rho_\alpha\}$ is a partition of unity subordinate to $\{V_\alpha\}$.

$$\therefore \mathcal{A}_p(p \cdot \xi) = \sum_\alpha \rho_\alpha(\pi(p))\mathcal{A}_{\alpha,p}(p \cdot \xi) \tag{55}$$

In this equation, $\rho_\alpha(\pi(p))$ is only non-zero on $\pi^{-1}(V_\alpha)$ and on each $\pi^{-1}(V_\alpha)$, $\mathcal{A}_{\alpha,p}(p \cdot \xi) = \xi$.
$\therefore \mathcal{A}_p(p \cdot \xi) = \sum_\alpha \rho_\alpha(\pi(p))\xi = \xi$, from the definition of a partition of unity.
For the other axiom, I can use the exact same logic about $\rho_\alpha(\pi(p))$ being non-zero only on $\pi^{-1}(V_\alpha)$, where the axiom holds, to get

$$(R_g)^*\mathcal{A} = \sum_\alpha (\rho_\alpha \circ \pi)(R_g)^*\mathcal{A}_\alpha = \sum_\alpha (\rho_\alpha \circ \pi)\mathrm{Ad}_{g^{-1}}\mathcal{A}_\alpha = \mathrm{Ad}_{g^{-1}}\mathcal{A} \tag{56}$$

as required. □

In physics, connections are introduced as compensating terms to make derivatives transform covariantly under some gauge redundancy. I will discuss this at length in section 2. But, before that, I will note a purely geometric interpretation of what the connection actually means on the total space[4].

Definition 1.17 (Vertical subspace)**.** *For a G-bundle, $\pi : P \to B$, the vertical subspace at $p \in P$, denoted $T^v_p P$, is defined to be $T^v_p P = \ker(\pi_* : T_p P \to T_{\pi(p)} B)$.*

Geometrically, the vertical subspace is the set of all directions within a single fibre embedded in P. Thus, $T^v_p P = T_p P_{\pi(p)}$ where $P_{\pi(p)} = \pi^{-1}(\pi(p))$ is the fibre. Since the canonical right G-action fully generates each fibre, it also follows that upon considering the canonical right $\mathfrak{g}$-action, $T^v_p P = p \cdot \mathfrak{g}$.

Definition 1.18 (Horizontal distribution)**.** *A horizontal subspace at p is defined to be any subspace whose direct sum with $T^v_p P$ is the whole of $T_p P$. A horizontal distribution is then defined to be a distribution on P (i.e. a subbundle of TP) which is a horizontal subspace at every point.*

The next theorem encapsulates the proclaimed purely geometric interpretation of a connection on a principal bundle.

Theorem 1.19. *A connection on a G-bundle, $\pi : P \to B$, is equivalent to choosing a horizontal distribution, H, satisfying $(R_g)^* H = H$, where R_g (again) denotes the canonical right G-action.*

Proof. First assume I have a connection, $\mathcal{A}$, on a G-bundle. Define a distribution, H, by $H = \ker(\mathcal{A})$.
$\mathcal{A}$ is $\mathfrak{g}$-valued and all values in $\mathfrak{g}$ can be output (as $\mathcal{A}_p(p \cdot \xi) = \xi$ and ξ is arbitrary).

$$\begin{aligned} \therefore \dim(H) &= \dim(P) - \dim(\mathfrak{g}) \ \text{ by the rank} - \text{nullity theorem} & (57)\\ &= \dim(P) - \dim(T^v P) \ \text{ as } \mathfrak{g} \text{ generates the vertical subspaces} & (58) \end{aligned}$$

[4]Typically, in the kind of discussion that follows, people end up making some reference or the other to parallel transport. I, on the other hand, will deliberately avoid any further mention of parallel transport because I don't think it's particularly useful and I don't understand it very well anyway.

meaning H has the correct dimension to be a horizontal distribution. Thus, I only have to show that it doesn't intersect any of the vertical subspaces non-trivially. For that, let $v \in H \cap T^v P$. At each point, p, elements of $T^v_p P$ can be written as $p \cdot \xi$ for some $\xi \in \mathfrak{g}$.
$\therefore \exists$ a ξ such that $p \cdot \xi = v_p$.
Then, $v_p \in H_p = \ker(\mathcal{A}_p) \implies 0 = \mathcal{A}_p(p \cdot \xi) = \xi \implies H$ really is a horizontal distribution.
For the right invariance condition, let $v \in H_p \iff \mathcal{A}_p(v) = 0$.
By definition, $(R_{g^{-1}})_* v \in T_{p \cdot g^{-1}} P$. Furthermore,

$$\mathcal{A}_{p \cdot g^{-1}}((R_{g^{-1}})_* v) = ((R_{g^{-1}})^* \mathcal{A}_{p \cdot g^{-1}})(v) \tag{59}$$
$$= \mathrm{Ad}_g \mathcal{A}_p(v) \text{ by the 2nd axiom of a connection} \tag{60}$$
$$= 0 \text{ as } \mathcal{A}_p(v) = 0 \tag{61}$$

$\therefore (R_{g^{-1}})_* v \in H_{p \cdot g^{-1}}$, which then implies $(R_{g^{-1}})_* H \subseteq H$.
$\therefore (R_g)^* H = H$ as $R_{g^{-1}}$ acts bijectively on each fibre and $R_g = (R_{g^{-1}})^{-1}$.
For the converse, assume I have a horizontal distribution with $(R_g)^* H = H$. Since a horizontal subspace direct sums with a vertical subspace to give the whole tangent space, it follows that every $v \in T_p P$ can be written uniquely as $v = p \cdot \xi + h$ for some $\xi \in \mathfrak{g}$ and $h \in H_p$.
Use this to define $\mathcal{A}$ as acting by $\mathcal{A}_p(v) = \mathcal{A}_p(p \cdot \xi + h) = \xi$.
Note that this definition is consistent with $H = \ker(\mathcal{A})$ from the first half of the proof. As for the axioms (of a connection) themselves, it's immediate that $\mathcal{A}_p(p \cdot \xi) = \xi$. The 2nd is less straightforward.

$$(R_g)^*(\mathcal{A}_p(p \cdot \xi + h)) = \mathcal{A}_{p \cdot g}((R_g)_*(p \cdot \xi + h)) \tag{62}$$
$$= \mathcal{A}_{p \cdot g}((R_g)_*(p \cdot \xi)) + \mathcal{A}_{p \cdot g}((R_g)_*(h)) \tag{63}$$

$(R_g)^* H = H \iff (R_{g^{-1}})_* H = H$ and g arbitrary means that $(R_g)_* h$ is still in the horizontal distribution.
$\therefore$ It lies in the kernel of $\mathcal{A}_{p \cdot g}$.

$$\therefore (R_g)^*(\mathcal{A}_p(p \cdot \xi + h)) = \mathcal{A}_{p \cdot g}((R_g)_*(p \cdot \xi)) \tag{64}$$
$$= \mathcal{A}_{p \cdot g}\left((R_g)_* \frac{\mathrm{d}}{\mathrm{d}t}\left(p \cdot \mathrm{e}^{t\xi}\right)\Big|_{t=0}\right) \tag{65}$$
$$= \mathcal{A}_{p \cdot g}\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(p \cdot \mathrm{e}^{t\xi} g\right)\Big|_{t=0}\right) \tag{66}$$
$$= \mathcal{A}_{p \cdot g}\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(p \cdot g g^{-1} \mathrm{e}^{t\xi} g\right)\Big|_{t=0}\right) \tag{67}$$
$$= \mathcal{A}_{p \cdot g}((p \cdot g) \cdot \mathrm{Ad}_{g^{-1}} \xi)) \tag{68}$$
$$= \mathrm{Ad}_{g^{-1}} \xi \tag{69}$$
$$= \mathrm{Ad}_{g^{-1}} \mathcal{A}_p(p \cdot \xi + h)) \tag{70}$$

Since p, ξ and h are arbitrary, it follows that $(R_g)^* \mathcal{A} = \mathrm{Ad}_{g^{-1}} \mathcal{A}$, thus proving that $\mathcal{A}$ really is a connection. ☐

Definition 1.20 (Horizontal section)**.** *A section (local or global), s, of a G-bundle, $\pi : P \to B$, with connection, $\mathcal{A}$, is called horizontal if and only if it's tangent to the associated horizontal distribution.*

Corollary 1.20.1. *s is horizontal if and only if $s^* \mathcal{A} = 0$.*

Proof. $\forall v \in T_{\pi(p)} B$, $(s^* \mathcal{A})(v) = \mathcal{A}(s_* v)$ and $H = \ker(\mathcal{A})$. ☐

The subject of connections is intimately linked to attempts to quantify the intuitive notions of (intrinsic) curvature. Philosphophically, curvature is the obstruction to flatness.

Definition 1.21 (Flat)**.** *A connection if called flat if and only if it's associated horizontal distribution is integrable (i.e. the distribution is closed under the Lie bracket, which is equivalent to the distribution arising from a foliation, due to Frobenius' theorem).*

There are a number of other ways to define flatness - I list them in the following theorem. From a physics point of view, the last one is the most useful and insightful.

Theorem 1.22. *For a connection, $\mathcal{A}$, on a G-bundle, $\pi : P \to B$, the following are equivalent.*

1. *$\mathcal{A}$ is flat.*
2. *P is foliated by local horizontal sections.*
3. *P has a local horizontal section over each point in B.*
4. *P can be covered by trivialisations, Φ_α, such that all the A_α are zero.*

Proof. (1) is equivalent to (2) from the definition of integrable.
(2) $\implies$ (3) because (2) is a stronger version of (3).
(3) $\iff$ (4) because trivialisations, Φ_α, correspond to local sections, s_α, which are horizontal if and only if $0 = (s_\alpha)^*\mathcal{A} = A_\alpha$.
Finally, (3) $\implies$ (2) as follows. Since (3) holds, For any point, $p \in P$, $\exists$ a section, s, over some open neighbourhood, V, of $\pi(p)$.
Then, the right translates, $(R_g)_*s$, foliate P over V and are horizontal by construction. ☐

Before defining curvature, I have to first define a confusing piece of notation.

Definition 1.23 (Bracket-wedge)**.** *Let σ and τ be $\mathfrak{g}$-valued differential forms. More explicitly, let $\sigma = \xi_i \otimes \sigma_i$ and $\tau = \eta_i \otimes \tau_i$, where ξ_i, $\eta_j \in \mathfrak{g}$, σ_i & τ_j are p & q forms respectively and the summation convention is in effect. Then, the bracket-wedge, $[\sigma \wedge \tau]$, is defined by*

$$[\sigma \wedge \tau] = [\xi_i, \eta_j] \otimes (\sigma_i \wedge \tau_j) \tag{71}$$

Corollary 1.23.1. *Because of the Lie bracket's and the wedge product's antisymmetry,* $[\sigma \wedge \tau] = (-1)^{pq+1}[\tau \wedge \sigma]$.

Definition 1.24 (Curvature)**.** *Given a connection, $\mathcal{A}$, on a G-bundle, $\pi : P \to B$, its curvature, $\mathcal{F}$, is a $\mathfrak{g}$-valued 2-form on P defined by $\mathcal{F} = \mathrm{d}\mathcal{A} + \frac{1}{2}[\mathcal{A} \wedge \mathcal{A}]$.*

I pontificated earlier that curvature was the obstruction to flatness. Hence, it is only natural that there must be some theorem like the next one.

Theorem 1.25. *A connection, $\mathcal{A}$, on a G-bundle, $\pi : P \to B$, is flat if and only if $\mathcal{F} = 0$.*

Proof. Let v be the vertical vector field defined by $v(p) = p \cdot \xi$ for some fixed $\xi \in \mathfrak{g}$.
By defintion, $\iota_v\mathcal{F} = \iota_v\mathrm{d}\mathcal{A} + \frac{1}{2}\iota_v[\mathcal{A} \wedge \mathcal{A}]$.
Decompose $\mathcal{A}$ as $\eta_i \otimes \sigma_i$.
$\therefore \frac{1}{2}\iota_v[\mathcal{A} \wedge \mathcal{A}] = \frac{1}{2}[\eta_i, \eta_j] \otimes \iota_v(\sigma_i \otimes \sigma_j - \sigma_j \otimes \sigma_i) = [\sigma_i(v)\eta_i, \eta_j] \otimes \sigma_j$.
But at each point, p, $\sigma_i(v)\eta_i = (\eta_i \otimes \sigma_i)(v) = \mathcal{A}_p(v(p)) = \mathcal{A}_p(p \cdot \xi) = \xi$.

$\therefore \iota_v[\mathcal{A}\wedge\mathcal{A}] = [\xi, \mathcal{A}]$, with the 1-form part of $\mathcal{A}$ just coming along for the ride in the Lie bracket. Then, from the definition of the Lie bracket,

$$[\xi, \mathcal{A}] = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{Ad}_{\mathrm{e}^{t\xi}}\mathcal{A}\Big)\Big|_{t=0} \tag{72}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}((R_{\mathrm{e}^{t\xi}})^*\mathcal{A})|_{t=0} \text{ by the 2nd connection axiom} \tag{73}$$

$$= -\mathcal{L}_v\mathcal{A} \text{ as } R_{\mathrm{e}^{t\xi}} \text{ is the flow of } -v \tag{74}$$

$$= -\iota_v\mathrm{d}\mathcal{A} - \mathrm{d}(\iota_v\mathcal{A}) \text{ by Cartan's magic formula} \tag{75}$$

But, as above, $\iota_v\mathcal{A} = \mathcal{A}(v) = \xi \implies \mathrm{d}(\iota_v\mathcal{A}) = \mathrm{d}\xi = 0$ as ξ doesn't vary across P. Hence, $[\xi, \mathcal{A}] = -\iota_v\mathrm{d}\mathcal{A}$.
$\therefore \iota_v\mathcal{F} = \iota_v\mathrm{d}\mathcal{A} + [\xi, \mathcal{A}] = 0 \implies \mathcal{F}(v, w) = 0$ for any vertical vector field, v[5].
$\therefore \mathcal{F}(v, w) = 0$ for any vertical vector field, w, as well, because $\mathcal{F}(v, w) = -\mathcal{F}(w, v)$.
$\therefore \mathcal{F} = 0 \iff \mathcal{F}(v, w) = 0$ when both v and w are horizontal vector fields.
Since the horizontal distribution is defined by $\ker(\mathcal{A})$, when acting on horizontal vector fields, $\mathcal{F}(v, w) = (\mathrm{d}\mathcal{A})(v, w)$.
Finally, by Frobenius' theorem, $\mathcal{A}$ is flat $\iff$ $\mathrm{d}\mathcal{A}$ vanishes when acting on $\ker(\mathcal{A}) = H$.
$\therefore \mathcal{F} = 0 \iff \mathcal{A}$ is flat. □

Like connections, in physics one tends to work with a local expression for curvature.

Definition 1.26 (Local curvature 2-form). *Given a section, s_α, corresponding to a trivialisation, Φ_α, the local curvature 2-form is defined to be $F_\alpha = (s_\alpha)^*\mathcal{F} = \mathrm{d}A_\alpha + \frac{1}{2}[A_\alpha \wedge A_\alpha]$.*

Corollary 1.26.1. *In local coordinates (strictly, this requires taking the intersection of the bundle trivialising patches and coordinate patches), dropping the α subscripts, the local curvature 2-form components, $F_{ij} = -F_{ji}$, are given by*

$$\frac{1}{2}F_{ij}\mathrm{d}x^i \wedge \mathrm{d}x^j = \mathrm{d}(A_i\mathrm{d}x^i) + \frac{1}{2}[A_i\mathrm{d}x^i, A_j\mathrm{d}x^j] \tag{76}$$

$$= \partial_j(A_i)\mathrm{d}x^j \wedge \mathrm{d}x^i + \frac{1}{2}[A_i, A_j]\mathrm{d}x^i \wedge \mathrm{d}x^j \tag{77}$$

$$\therefore F_{ij}\mathrm{d}x^i \otimes \mathrm{d}x^j = (\partial_i(A_j) - \partial_j(A_i) + [A_i, A_j])\mathrm{d}x^i \otimes \mathrm{d}x^j \tag{78}$$

$$\therefore F_{ij} = \partial_iA_j - \partial_jA_i + [A_i, A_j] \tag{79}$$

This famous formula will rear its head repeatedly in section 2. Unlike connections, the local connection 2-forms also transform much more nicely on overlaps.

Theorem 1.27. *On trivialising patch overlaps, $V_\alpha \cap V_\beta$, the local connection 2-forms transform as $F_\beta = \mathrm{Ad}_{g_{\beta\alpha}}F_\alpha$.*

Proof. Using theorem 1.14,

$$F_\beta = \mathrm{d}A_\beta + \frac{1}{2}[A_\beta \wedge A_\beta] \tag{80}$$

$$= \mathrm{d}\Big(g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} + \mathrm{Ad}_{g_{\beta\alpha}}A_\alpha\Big) + \frac{1}{2}\Big[\Big(g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} + \mathrm{Ad}_{g_{\beta\alpha}}A_\alpha\Big) \wedge \Big(g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} + \mathrm{Ad}_{g_{\beta\alpha}}A_\alpha\Big)\Big] \tag{81}$$

In this case, I will write $\mathrm{Ad}_{g_{\beta\alpha}}A_\alpha$ as $g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1}$ as convenient notation (with transtions functions on the right and left being understood as pushforwards under right and left multiplication

[5] I think there may be an issue here about whether ξ can really be taken as a constant in the definition of v.

respectively[6]). Then,

$$\begin{aligned}F_\beta &= \mathrm{d}(g_{\beta\alpha}) \wedge \mathrm{d}g_{\beta\alpha}^{-1} + 0 + \mathrm{d}(g_{\beta\alpha}) \wedge A_\alpha g_{\beta\alpha}^{-1} + g_{\beta\alpha}\mathrm{d}(A_\alpha)g_{\beta\alpha}^{-1} - g_{\beta\alpha}A_\alpha \wedge \mathrm{d}g_{\beta\alpha}^{-1} \\ &+ \frac{1}{2}[g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}] + \frac{1}{2}[g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}] + \frac{1}{2}[g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1}] \\ &+ \frac{1}{2}[g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1}] \end{aligned} \tag{82}$$

The $\frac{1}{2}$ factors accommodate for the paired of the Lie bracket and the wedge product. Thus,

$$\frac{1}{2}[g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}] = g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}g_{\beta\alpha} \wedge \mathrm{d}g_{\beta\alpha}^{-1} = -\mathrm{d}(g_{\beta\alpha}) \wedge \mathrm{d}g_{\beta\alpha}^{-1} \tag{83}$$

$$\frac{1}{2}[g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}] = g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1}g_{\beta\alpha} \wedge \mathrm{d}g_{\beta\alpha}^{-1} = g_{\beta\alpha}A_\alpha \wedge \mathrm{d}g_{\beta\alpha}^{-1} \tag{84}$$

$$\frac{1}{2}[g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}A_\alpha] = g_{\beta\alpha}\mathrm{d}g_{\beta\alpha}^{-1}g_{\beta\alpha} \wedge A_\alpha = -\mathrm{d}(g_{\beta\alpha}) \wedge A_\alpha \tag{85}$$

$$\frac{1}{2}[g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1} \wedge g_{\beta\alpha}A_\alpha g_{\beta\alpha}^{-1}] = \frac{1}{2}g_{\beta\alpha}[A_\alpha \wedge A_\alpha]g_{\beta\alpha}^{-1} \tag{86}$$

Inserting these four expressions into equation 82 leaves $F_\beta = g_{\beta\alpha}(\mathrm{d}A_\alpha + \frac{1}{2}[A_\alpha \wedge A_\alpha])g_{\beta\alpha}^{-1}$, which is exactly $\mathrm{Ad}_{g_{\beta\alpha}}F_\alpha$. □

I have now devoted many pages to connections on principal bundles. Luckily, connections on vector bundles are not a huge leap from here - indeed a connection on a vector bundle will always be derived from one on a principal bundle.

Definition 1.28 (Associated vector bundle)**.** *Let $\rho : G \to \mathrm{GL}(V)$ be a (linear) representation of G and let $\pi : P \to B$ be a G-bundle. Then, the associated vector bundle is defined to be $(P \times V)/((p \cdot g, v) \sim (p, \rho(g)v))$. This is a vector bundle over B and is denoted $P \times_G V$. In more elementary terms, for a G-bundle with transition functions, $g_{\beta\alpha}$, the associated vector bundle is the vector bundle with transition functions, $\rho(g_{\beta\alpha})$.*

Proof. For the proof, the equivalence relation, $(p \cdot h, v) \sim (p, \rho(h)v)$, is more conveniently written as $(p, v) \sim (p \cdot h, \rho(h)^{-1}v)$.
Let $\Phi_\alpha : \pi^{-1}(V_\alpha) \to V_\alpha \times G$ be local trivialisations of P. Then, $\Phi_\alpha(p) = (b, g)$ for some $b \in V_\alpha$ and $g \in G$.
Then, define local trivialisations of $P \times_G V$, $\Phi'_\alpha : \pi^{-1}(V_\alpha) \to V_\alpha \times V$, by $\Phi'_\alpha(p, v) = (b, \rho(g)v)$. This is well defined because the equivalent point, $(p \cdot h, \rho(h)^{-1}v)$, has $\Phi_\alpha(p \cdot h) = (b, gh)$ by definition and thus $\Phi'_\alpha(p \cdot h, \rho(h)^{-1}v) = (b, \rho(gh)\rho(h)^{-1}v) = (b, \rho(g)v) = \Phi'_\alpha(p, v)$.
Next, consider p on overlaps. On P, the transition functions are defined so that
$\Phi_\alpha(p) = (b, g) \implies \Phi_\beta(p) = (b, g_{\beta\alpha}(p)g)$.
$\therefore (\Phi'_\beta \circ \Phi'_\alpha)(b, v) = \Phi'_\beta(p, \rho(g)^{-1}v) = (b, \rho(g_{\beta\alpha}(p)g)\rho(g)^{-1}v) = (b, \rho(g_{\beta\alpha}(p))v)$.
$\therefore$ The transition functions of $P \times_G V$ are indeed $\rho(g_{\beta\alpha})$. □

In some sense, the association goes the other way too. In particular, for every vector bundle, there is one particular principal bundle that will play a useful role in what follows.

Definition 1.29 (Frame bundle)**.** *Let $\pi : E \to B$ be a rank-k vector bundle with transition functions, $g_{\beta\alpha} \in \mathrm{GL}(k, \mathbb{R})$. Let $F(\mathbb{R}^k)$ be the set of ordered bases in $\mathbb{R}^k$. Then, the frame bundle*

[6] I believe this interpretation is fine for what follows, although I am sometimes suspicious the rest of the proof only works for matrix Lie groups.

of E, denoted $F(E)$, is a principal $\mathrm{GL}(k,\mathbb{R})$*-bundle defined to be the space of ordered bases in each fibre, i.e.*

$$F(E) = \left(\bigsqcup_{\alpha} V_\alpha \times F(\mathbb{R}^k)\right) \Big/ \left(\left(b \in V_\alpha, (e_1, \ldots, e_k)\right) \sim \left(b \in V_\beta, (g_{\beta\alpha}(b)e_1, \ldots, g_{\beta\alpha}(b)e_k)\right)\right) \tag{87}$$

$\therefore$ *By construction, $F(E)$ is a principal bundle with the same transition functions as the vector bundle, E.*

Corollary 1.29.1. *If $\rho : \mathrm{GL}(k,\mathbb{R}) \to \mathrm{GL}(\mathbb{R}^k)$ is the defining representation, then the associated vector bundle of $F(E)$ is E itself.*

Definition 1.30 (Connection on a vector bundle)**.** *A connection on a vector bundle is defined to be a connection on its frame bundle.*

Definition 1.31 (Induced connection)**.** *Given a connection, $\mathcal{A}$, on a G-bundle, $\pi : P \to B$, and a representation, $\rho : G \to \mathrm{GL}(V)$, the induced connection on the associated vector bundle is defined by local connection 1-forms, $\rho_* A_\alpha$ (where ρ_* acts on the $\mathfrak{g}$-valued part of A_α).*

Having gone through this seemingly never-ending jiggery-pokery, I have finally completed laying the foundations for the main purpose of connections - differentiating vectors on curved spaces.

Definition 1.32 (Covariant derivative on a vector bundle)**.** *Let $\pi : E \to B$ be a rank-k vector bundle, let $\mathcal{A}$ be a connection on E, let $s : B \to E$ be a section and let $pr_2 \circ \Phi_\alpha(s_{|V_\alpha}) = v_\alpha \in \mathbb{R}^k$ be the local version of s. Then, the covariant derivative of s with respect to $\mathcal{A}$, denoted $\mathrm{d}^{\mathcal{A}}s$, is the E-valued 1-form on B defined locally by $\mathrm{d}v_\alpha + A_\alpha v_\alpha$.*

Proof. It has to be checked that these local constructions glue together to give a well defined global object.
Note that because $\mathrm{GL}(k,\mathbb{R})$ is a matrix Lie group, pushforwards are just matrix multiplication. Furthermore, $\mathfrak{gl}(\mathfrak{k},\mathbb{R})$ is just the set of all $k \times k$ matrices.
In a different trivialising patch, $v_\beta = g_{\beta\alpha} v_\alpha$ by the definition of transition functions and $A_\beta = g_{\beta\alpha} A_\alpha g_{\beta\alpha}^{-1} + g_{\beta\alpha} \mathrm{d} g_{\beta\alpha}^{-1}$ by theorem 1.14.

$$\begin{aligned}
\mathrm{d}v_\beta + A_\beta v_\beta &= \mathrm{d}(g_{\beta\alpha} v_\alpha) + (g_{\beta\alpha} A_\alpha g_{\beta\alpha}^{-1} + g_{\beta\alpha}\mathrm{d}(g_{\beta\alpha}^{-1})) g_{\beta\alpha} v_\alpha && (88)\\
&= \mathrm{d}(g_{\beta\alpha}) v_\alpha + g_{\beta\alpha}\mathrm{d}v_\alpha + g_{\beta\alpha} A_\alpha v_\alpha + g_{\beta\alpha}\mathrm{d}(g_{\beta\alpha}^{-1}) g_{\beta\alpha} v_\alpha && (89)\\
&= \mathrm{d}(g_{\beta\alpha}) v_\alpha + g_{\beta\alpha}\mathrm{d}v_\alpha + g_{\beta\alpha} A_\alpha v_\alpha - \mathrm{d}(g_{\beta\alpha}) v_\alpha && (90)\\
&= g_{\beta\alpha}(\mathrm{d}v_\alpha + A_\alpha v_\alpha) && (91)
\end{aligned}$$

which is the required transformation property for $\mathrm{d}v_\alpha + A_\alpha v_\alpha$ to define a globally E-valued object. The 1-form part requires no checking; it is manifest already. □

This proof illustrates the reason the A_α are thought of as compensating fields in physics. They are non-homogeneous, non-tensorial additions to $\mathrm{d}v_\alpha$ - the "naive" derivative - inserted to make the resultant object transform properly under a change of local trivialisation.

These definitions also have a little bit more generality than might initially appear. From any representation, $\rho : G \to \mathrm{GL}(V)$, I can construction its dual representation and subsequently various tensor products of representations. This way, given a covariant derivative on a vector bundle, $\pi : E \to B$, I also get covariant derivatives of sections of $E^*, E \otimes E^*$ etc. I'll give concrete examples of how this works in section 2.1.

There is also the following generalisation to the covariant derivative just defined.

Definition 1.33 (Exterior covariant derivative)**.** *Let $\pi : E \to B$ be a vector bundle and let σ be an E-valued p-form on B. Therefore, locally σ can be written as the sum of expressions, $s \otimes \alpha$, where s is a section of E and α is a p-form.*
The exterior covariant derivative, $\mathrm{d}^{\mathcal{A}}\sigma$, is then defined locally by $\mathrm{d}^{\mathcal{A}}(s\otimes\alpha) = (\mathrm{d}^{\mathcal{A}}s)\wedge\alpha + s\otimes\mathrm{d}\alpha$.

The notation is a bit clunky here, but its application in the next theorem will hopefully make the meaning clearer.

Theorem 1.34. $(\mathrm{d}^{\mathcal{A}})^2\sigma = F \wedge \sigma$*, where F is interpreted as a $\mathfrak{gl}(V)$-valued 2-form acting on the E-valued part of σ and E's fibres are the vector space, V.*

Proof. Let σ be $s \otimes w$ locally and let the section, s, be v_α locally. Then,

$$\mathrm{d}^{\mathcal{A}}(v_\alpha \otimes w) = \mathrm{d}^{\mathcal{A}}(v_\alpha) \wedge w + v_\alpha \otimes \mathrm{d}w \tag{92}$$

$$= (\mathrm{d}v_\alpha) \wedge w + (A_\alpha v_\alpha) \wedge w + v_\alpha \otimes \mathrm{d}w \tag{93}$$

where in the first two terms (and henceforth in this proof) I'm adopting the convention of leaving off the $\otimes$ when there's already an infix symbol, $\wedge$ in this case.
The rest of the proof is easiest in local coordinates.
$\therefore \mathrm{d}v_\alpha = \partial_\mu(v_i)e_i \otimes \mathrm{d}x^\mu$ where $\{e_i\}_{i=1}^k$ is some basis of $\mathbb{R}^k$, $A_\alpha = A_{ij\mu}\mathrm{d}x^\mu$ where the i and j are the matix indices of elements of $\mathfrak{gl}(\mathfrak{k}, \mathbb{R})$ and $w = w_I\mathrm{d}x^I$ where $\mathrm{d}x^I$ is multi-index notation.

$$\therefore \mathrm{d}^{\mathcal{A}}(v_\alpha \otimes w) = (\partial_\mu(v_i)w_I + A_{ij\mu}v_jw_I + v_i\partial_\mu(w_I))e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^I \tag{94}$$

I'll apply the second $\mathrm{d}^{\mathcal{A}}$ separately to each term on the RHS. Omitting terms that are zero due to symmetries being contracted with antisymmetries, I get the following.

$$\mathrm{d}^{\mathcal{A}}(\partial_\mu(v_i)w_Ie_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^I)$$

$$= \partial_\nu(\partial_\mu(v_i)w_I)e_i \otimes \mathrm{d}x^\nu \wedge \mathrm{d}x^\mu \wedge \mathrm{d}x^I + A_{ij\nu}\partial_\mu(v_j)w_Ie_i \otimes \mathrm{d}x^\nu \wedge \mathrm{d}x^\mu \wedge \mathrm{d}x^I \tag{95}$$

$$= (A_{ij\mu}\partial_\nu(v_j)w_I + \partial_\nu(v_i)\partial_\mu(w_I))e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{96}$$

$$\mathrm{d}^{\mathcal{A}}(A_{ij\mu}v_jw_Ie_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^I)$$

$$= (\partial_\mu(A_{ij\nu}v_jw_I) + A_{ik\mu}A_{kj\nu}v_jw_I)e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{97}$$

$$\mathrm{d}^{\mathcal{A}}(v_i\partial_\mu(w_I)e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^I) \tag{98}$$

$$= (\partial_\mu(v_i)\partial_\nu(w_I) + A_{ij\mu}v_j\partial_\nu(w_I))e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{99}$$

Putting them all together,

$$\begin{aligned}(\mathrm{d}^{\mathcal{A}})^2\sigma = (&A_{ij\mu}\partial_\nu(v_j)w_I + \partial_\nu(v_i)\partial_\mu(w_I) + \partial_\mu(A_{ij\nu}v_jw_I) + A_{ik\mu}A_{kj\nu}v_jw_I + \partial_\mu(v_i)\partial_\nu(w_I) \\ &+ A_{ij\mu}v_j\partial_\nu(w_I))e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \end{aligned}\tag{100}$$

$$\begin{aligned}= (&A_{ij\mu}\partial_\nu(v_j)w_I + \partial_\nu(v_i)\partial_\mu(w_I) + \partial_\mu(A_{ij\nu})v_jw_I + A_{ij\nu}\partial_\mu(v_j)w_I + A_{ij\nu}v_j\partial_\mu(w_I) \\ &+ A_{ik\mu}A_{kj\nu}v_jw_I + \partial_\mu(v_i)\partial_\nu(w_I) + A_{ij\mu}v_j\partial_\nu(w_I))e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I\end{aligned} \tag{101}$$

$$= (\partial_\mu(A_{ij\nu})v_jw_I + A_{ik\mu}A_{kj\nu}v_jw_I)e_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \text{ by } \mu - \nu \text{ antisymmetry} \tag{102}$$

$$= (\partial_\mu(A_{ij\nu}) + A_{ik\mu}A_{kj\nu})v_jw_Ie_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{103}$$

On the other hand, $F_{\mu\nu} = \partial_\mu(A_\nu) - \partial_\nu(A_\mu) + [A_\mu, A_\nu]$ from equation 79. Restoring the i, j indices and noting that the Lie bracket in $\mathfrak{gl}(\mathfrak{k}, \mathbb{R})$ is just the commutator,

$$F \wedge \sigma = \left(\frac{1}{2}F_{ij\mu\nu}v_je_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu\right) \wedge (w_I\mathrm{d}x^I) \tag{104}$$

$$= \frac{1}{2}(\partial_\mu(A_{ij\nu}) - \partial_\nu(A_{ij\mu}) + A_{ik\mu}A_{kj\nu} - A_{ik\nu}A_{kj\mu})v_jw_Ie_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{105}$$

$$= (\partial_\mu(A_{ij\nu}) + A_{ik\mu}A_{kj\nu})v_jw_Ie_i \otimes \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \wedge \mathrm{d}x^I \tag{106}$$

which is exactly $(\mathrm{d}^{\mathcal{A}})^2\sigma$ from above. □

2 Applications

2.1 Riemannian geometry

The simplest application of the general theory discussed in section 1 is the construction of covariant derivatives on the tangent bundle. These are exactly the standard covariant derivatives of tensor fields that is one is introduced to in a standard exposition of Riemannian geometry. In fact, it's to make connection[7] with these familiar results that I've chosen to present this application first.

Definition 2.1 (Connection on a manifold)**.** *For any manifold, M, a connection on M is defined to be connection on its tangent bundle, TM.*

Definition 2.2 (∇ notation)**.** *In this context, the covariant derivative is denoted ∇. Furthermore, the one form component gained when acting with ∇ is denoted with a subscript like ∇_i. For example, given a vector field, $v = v^i(x)\partial_i$ in local coordinates, it's covariant derivative, ∇v, is denoted $\nabla v = \nabla_i(v^j)\mathrm{d}x^i \otimes \partial_j$ in local coordinates.*

Throughout this section, I'll be using local coordinate patches to trivialise TM.
$\therefore$ The transition functions are just change of coordinate matrices, $\partial x'^i/\partial x^j$.
Furthermore, these matrices must be invertible because change of coordinates must be invertible.
$\therefore$ The structure group of TM is $\mathrm{GL}(n, \mathbb{R})$, where $n = \dim(M)$ throughout.

Definition 2.3 (Christoffel symbols)**.** *The ($\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$-valued) local connection 1-forms on a manifold are called Christoffel symbols and denoted ${\Gamma^i}_{jk}(x)\mathrm{d}x^k$ in local coordinates.*

Given the local connection 1-forms, A_α, act by matrix multiplication in definition 1.32, the positioning of the i and j indices in ${\Gamma^i}_{jk}(x)\mathrm{d}x^k$ is meant to indicate that being ($\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$-valued) at each point means the i and j indices form a local section of the endomorphism bundle, $\mathrm{End}(TM)$.

Theorem 2.4. *Under a change of coordinates, $x^i \mapsto x'^i$, the Christoffel symbols transform as*

$$ {\Gamma'^i}_{jk} = \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}{\Gamma^l}_{mn} + \frac{\partial x'^i}{\partial x^l}\frac{\partial^2 x^l}{\partial x'^j x'^k} \tag{107} $$

Proof. Proving this theorem is only a matter of applying theorem 1.14 to this particular instance of a connection. Here, $\mathfrak{g} = \mathfrak{gl}(\mathfrak{n}, \mathbb{R})$ and thus like with any matrix Lie group, all pushforwards are just matrix multiplication. Furthermore, since I'm using local coordinate patches to trivialise the tangent bundle,

$$ g_{\beta\alpha}(p) = \left.\frac{\partial x'}{\partial x}\right|_p \text{ and } g^{-1}_{\beta\alpha}(p) = \left.\frac{\partial x}{\partial x'}\right|_p \tag{108} $$

Substituting these into theorem 1.14,

$$ {\Gamma'^i}_{jk}\mathrm{d}x'^k = \frac{\partial x'^i}{\partial x^l}{\Gamma^l}_{mk}\frac{\partial x^m}{\partial x'^j}\mathrm{d}x^k + \frac{\partial x'^i}{\partial x^l}\mathrm{d}\left(\frac{\partial x^l}{\partial x'^j}\right) \tag{109} $$

$$ = \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}{\Gamma^l}_{mn}\mathrm{d}x'^k + \frac{\partial x'^i}{\partial x^l}\frac{\partial^2 x^l}{\partial x'^k x'^j}\mathrm{d}x'^k \tag{110} $$

$\therefore$ From the coefficients of $\mathrm{d}x'^k$ it follows that the Christoffel symbols transform exactly as claimed in the theorem. □

[7] no pun intended

Theorem 2.5. *In local coordinates, the covariant derivative acts on an arbitrary tensor as*

$$\nabla_i T^{j_1\cdots j_a}{}_{k_1\cdots k_b} = \partial_i T^{j_1\cdots j_a}{}_{k_1\cdots k_b} + \sum_{l=1}^{a} \Gamma^{j_l}{}_{mi} T^{j_1\cdots \hat{j_l} m\cdots j_a}{}_{k_1\cdots k_b} - \sum_{l=1}^{b} \Gamma^{m}{}_{k_l i} T^{j_1\cdots j_a}{}_{k_1\cdots \hat{k_l} m\cdots k_b} \tag{111}$$

where a hat denotes an omitted index in a list.

Proof. Given a representation, $\rho : \mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(V)$, the induced connection is ρ_* of the original connection.
Vectors transform under the defining representation. Hence, by definition 1.32, if $v = v^i\partial_i$ in local coordinates, then

$$\nabla v^i = \mathrm{d}v^i + (\Gamma \cdot v)^i \tag{112}$$

$$= \partial_j(v^i)\mathrm{d}x^j + \Gamma^i{}_{kj} v^k \mathrm{d}x^j \tag{113}$$

$$\therefore \nabla_j v^i = \partial_j(v^i) + \Gamma^i{}_{kj} v^k \tag{114}$$

From here, I can build up the general result based on the various tensor representations of $\mathrm{GL}(n,\mathbb{R})$. In particular, covectors transform under the dual of the defining representation, $\rho(A) = A^{-T}$ for any $A \in \mathrm{GL}(n,\mathbb{R})$.
$\therefore$ The general tensor representation is $\rho(A) = A \otimes \cdots \otimes A \otimes A^{-T} \otimes \cdots \otimes A^{-T}$, with a lots of A and b lots of A^{-T}.
Since $\mathfrak{g} = T_eG$, $\Gamma \in \mathfrak{gl}(\mathfrak{n},\mathbb{R}) \otimes T^*M$ can be viewed (for the $\mathfrak{gl}(\mathfrak{n},\mathbb{R})$ part) as the equivalence class, $[\mathrm{e}^{t\Gamma}]$.
Hence, the connection I need for the general tensor product is

$$\rho_*[\mathrm{e}^{t\Gamma}] = [\rho(\mathrm{e}^{t\Gamma})] \tag{115}$$

$$= [\mathrm{e}^{t\Gamma} \otimes \cdots \otimes \mathrm{e}^{t\Gamma} \otimes \mathrm{e}^{-t\Gamma^T} \otimes \cdots \otimes \mathrm{e}^{-t\Gamma^T}] \tag{116}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{t\Gamma} \otimes \cdots \otimes \mathrm{e}^{t\Gamma} \otimes \mathrm{e}^{-t\Gamma^T} \otimes \cdots \otimes \mathrm{e}^{-t\Gamma^T})|_{t=0} \tag{117}$$

$$= \sum_{i=1}^{a} \Gamma \text{ in } i^{\text{th}} \text{ place} - \sum_{i=1}^{b} \Gamma^T \text{ in } (a+i)^{\text{th}} \text{ place} \tag{118}$$

with I in all the remaining tensor product slots in the sum of the last line.
With the Γ or Γ^T now acting on the corresponding index of $T^{j_1\cdots j_a}{}_{k_1\cdots k_b}$, I get exactly the result claimed in the statement of the theorem. □

Corollary 2.5.1. *The covariant derivative satisfies the Liebniz rule.*

Proof. Using the theorem,

$$\begin{aligned}
&\nabla_i(S^{j_1\cdots j_a}{}_{k_1\cdots k_b} T^{l_1\cdots l_c}{}_{m_1\cdots m_d}) \\
&= \partial_i(S^{j_1\cdots j_a}{}_{k_1\cdots k_b} T^{l_1\cdots l_c}{}_{m_1\cdots m_d}) + \sum_{n=1}^{a} \Gamma^{j_n}{}_{pi} S^{j_1\cdots \hat{j_n} p\cdots j_a}{}_{k_1\cdots k_b} T^{l_1\cdots l_c}{}_{m_1\cdots m_d} \\
&\quad + \sum_{n=1}^{c} \Gamma^{l_n}{}_{pi} S^{j_1\cdots j_a}{}_{k_1\cdots k_b} T^{l_1\cdots \hat{l_n} p\cdots l_c}{}_{m_1\cdots m_d} - \sum_{n=1}^{b} \Gamma^{p}{}_{k_n i} S^{j_1\cdots j_a}{}_{k_1\cdots \hat{k_n} p\cdots k_b} T^{l_1\cdots l_c}{}_{m_1\cdots m_d} \\
&\quad - \sum_{n=1}^{d} \Gamma^{p}{}_{m_n i} S^{j_1\cdots j_a}{}_{k_1\cdots k_b} T^{l_1\cdots l_c}{}_{m_1\cdots \hat{m_n} p\cdots m_d}
\end{aligned} \tag{119}$$

$$\therefore \nabla_i(S^{j_1\cdots j_a}{}_{k_1\cdots k_b}T^{l_1\cdots l_c}{}_{m_1\cdots m_d})$$

$$= \left(\partial_i(S^{j_1\cdots j_a}{}_{k_1\cdots k_b}) + \sum_{n=1}^{a}\Gamma^{j_n}{}_{pi}S^{j_1\cdots \hat{j_n}p\cdots j_a}{}_{k_1\cdots k_b} - \sum_{n=1}^{b}\Gamma^{p}{}_{k_n i}S^{j_1\cdots j_a}{}_{k_1\cdots \hat{k_n}p\cdots k_b}\right)T^{l_1\cdots l_c}{}_{m_1\cdots m_d}$$

$$+ S^{j_1\cdots j_a}{}_{k_1\cdots k_b}\left(\partial_i(T^{l_1\cdots l_c}{}_{m_1\cdots m_d}) + \sum_{n=1}^{c}\Gamma^{l_n}{}_{pi}T^{l_1\cdots \hat{l_n}p\cdots l_c}{}_{m_1\cdots m_d} - \sum_{n=1}^{d}\Gamma^{p}{}_{m_n i}T^{l_1\cdots l_c}{}_{m_1\cdots \hat{m_n}p\cdots m_d}\right) \tag{120}$$

$$= \nabla_i(S^{j_1\cdots j_a}{}_{k_1\cdots k_b})T^{l_1\cdots l_c}{}_{m_1\cdots m_d} + S^{j_1\cdots j_a}{}_{k_1\cdots k_b}\nabla_i(T^{l_1\cdots l_c}{}_{m_1\cdots m_d}) \tag{121}$$

which is exactly the Liebniz rule □

Definition 2.6 (Torsion)**.** *Let θ be the TM valued 1-form, $\theta(v) = v$ for any vector field, v. The torsion of a connection is defined to be $T = \mathrm{d}^{\mathcal{A}}\theta$. A connection is called torsion-free if and only if $T = 0$.*

Theorem 2.7. *T is a tensor that in local coordinates is $T = -(\Gamma^i{}_{jk} - \Gamma^i{}_{kj})\partial_i \otimes \mathrm{d}x^j \otimes \mathrm{d}x^k$.*

Proof. $\theta(\partial_i) = \partial_i = \delta^j{}_i\partial_j$.
$\therefore$ In components, θ is $\delta^j{}_i$, where j is the "TM-valued" part and i is the "1-form" part. Then, from definition 1.33,

$$T = \mathrm{d}^{\mathcal{A}}\theta \tag{122}$$
$$= \mathrm{d}^{\mathcal{A}}(\delta^i{}_j\partial_i\mathrm{d}x^j) \tag{123}$$
$$= (\partial_k(\delta^i{}_j) + \Gamma^i{}_{lk}\delta^l{}_j)\partial_i \otimes \mathrm{d}x^k \wedge \mathrm{d}x^j + \delta^i{}_j\partial_i \otimes \mathrm{d}^2x^j \tag{124}$$
$$= -(\Gamma^i{}_{jk} - \Gamma^i{}_{kj})\partial_i \otimes \mathrm{d}x^j \otimes \mathrm{d}x^k \tag{125}$$

as claimed. □

Corollary 2.7.1. *A torsion-free connection has $\Gamma^i{}_{jk} = \Gamma^i{}_{kj}$ in local coordinates.*

Corollary 2.7.2. *In a coordinate-free way, the torsion tensor acts on arbitrary vector fields, v and w, by $T(v, w) = \nabla_v w - \nabla_w v - [v, w]$, where $\nabla_v = v^a\nabla_a$.*

Proof. In local coordinates,

$$\text{RHS} = v^j\nabla_j w^i - w^j\nabla_j v^i - v^j\partial_j w^i + w^j\partial_j v^i \tag{126}$$
$$= v^j\partial_j w^i + v^j\Gamma^i{}_{kj}w^k - w^j\partial_j v^i - w^j\Gamma^i{}_{kj}v^k - v^j\partial_j w^i + w^j\partial_j v^i \tag{127}$$
$$= -(\Gamma^i{}_{jk} - \Gamma^i{}_{kj})v^j w^k \tag{128}$$
$$= \text{LHS} \tag{129}$$

Since both sides are tensors, agreeing in one basis means they agree in general. □

Definition 2.8 (Metric on a vector bundle)**.** *A metric, g, on a vector bundle, $\pi : E \to B$, is a section of $E^* \otimes E^*$ that is fibrewise symmetric and non-degenerate.*

In more down-to-Earth terms, a metric is a bilinear form on each fibre that fits together in some smooth way. Unfortunately, it has nothing directly to do with metric spaces. I have always felt the name was simply chosen by a sadist who sought to confuse students for all posterity.

Definition 2.9 (Inner product on a vector bundle)**.** *If a metric on a vector bundle is fibrewise positive-definite, then it is called an inner product.*

Definition 2.10 (Metrics on a manifold)**.** *A metric on a manifold is defined to be a metric on the tangent bundle. If the metric is positive-definite, it is called Riemannian.*

Definition 2.11 (Raising and lower indices)**.** *Given a metric, g, on a manifold, let g^{ij} denote the components of the inverse matrix of g_{ij}. Thus, g^{ij} defines a type-(2, 0) tensor (because non-degenerate bilinear forms define isomorphisms between a vector space and its dual vector space). Furthermore, indices are raised and lowered using g^{ij} and g_{ij} respectively.*

For example, given an object, V_i, on a manifold, V^i would mean $g^{ij}V_j$. Likewise, given an object, W^i, on an manifold, W_i would mean $g_{ij}W^j$.

Theorem 2.12. *Every vector bundle, $\pi : E \to B$, admits an inner product.*

Proof. Cover E with trivialisations, $\Phi_\alpha^{-1} : \pi^{-1}(V_\alpha) \to V_\alpha \times \mathbb{R}^k$.
On each $\pi^{-1}(V_\alpha)$, there's an inner product, g_α, corresponding to the standard inner product on $\mathbb{R}^k$.
Let $\{\rho_\alpha\}$ be a partition of unity subordinate to $\{V_\alpha\}$.
Finally, let $g = \sum_\alpha \rho_\alpha g_\alpha$. As each g_α is bilinear, symmetric and positive-definite, so is g. □

Corollary 2.12.1. *Every manifold admits a Riemannian metric.*

Observe that the theorem's proof relies critically on positive-definiteness. If any other metric signature were used, then upon summing in the partition of unity, the result could be anything. Unfortunately, in physics, metrics are almost always needed in the context of relativity, where the metric must be Lorentzian, i.e. a bilinear form with signature, $(-1, 1, \cdots, 1)$, which is not positive-definite. Indeed, I am reliably informed that not every manifold admits a metric of this sort.

Definition 2.13 (Metric-compatible)**.** *A connection on a manifold with metric, g, is called metric-compatible if and only if $\nabla g = 0$.*

Definition 2.14 (Levi-Civita connection)**.** *On any manifold with metric, there exists a unique metric-compatible, torsion-free connection, called the Levi-Civita connection.*

Proof. It suffices to just find the Christoffel symbols for the Levi-Civita connection. From corollary 2.7.1, ${\Gamma^i}_{jk} = {\Gamma^i}_{kj}$.
Then, from theorem 2.5, the metric compatibility condition says

$$0 = \partial_i g_{jk} - {\Gamma^l}_{ji} g_{lk} - {\Gamma^l}_{ki} g_{jl} \tag{130}$$
$$= \partial_i g_{jk} - \Gamma_{kji} - \Gamma_{jki} \tag{131}$$

Then, with the benefit of already knowing the correct answer, observe that the previous equation means

$$\frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) = \frac{1}{2}(\Gamma_{kji} + \Gamma_{jki} + \Gamma_{ikj} + \Gamma_{kij} - \Gamma_{jik} - \Gamma_{ijk}) \tag{132}$$
$$= \Gamma_{kij} \quad \text{using the torsion} - \text{free condition} \tag{133}$$

∴ The connection components are uniquely determined.
To prove existence though, I still need to check that these connection components satisfy the transformation law of theorem 2.4. Satisfying that transformation law guarantees I have a well-defined connection because of theorem 1.15

The metric and its inverse are tensors, so their components transform as per the standard rules for tensors. Hence,

$$\Gamma'^i{}_{jk} = \frac{1}{2} g'^{il}(\partial'_j g'_{kl} + \partial'_k g'_{lj} - \partial'_l g'_{jk}) \tag{134}$$

$$= \frac{1}{2}\frac{\partial x'^i}{\partial x^m}\frac{\partial x'^l}{\partial x^n} g^{mn}\left(\frac{\partial x^p}{\partial x'^j}\partial_p\left(\frac{\partial x^q}{\partial x'^k}\frac{\partial x^r}{\partial x'^l}g_{qr}\right) + \frac{\partial x^p}{\partial x'^k}\partial_p\left(\frac{\partial x^q}{\partial x'^l}\frac{\partial x^r}{\partial x'^j}g_{qr}\right) - \frac{\partial x^p}{\partial x'^l}\partial_p\left(\frac{\partial x^q}{\partial x'^j}\frac{\partial x^r}{\partial x'^k}g_{qr}\right)\right) \tag{135}$$

$$= \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}\Gamma^l{}_{mn} + \frac{1}{2}\frac{\partial x'^i}{\partial x^m}\frac{\partial x'^l}{\partial x^n} g^{mn}\left(g_{qr}\frac{\partial x^r}{\partial x'^l}\frac{\partial^2 x^q}{\partial x'^j x'^k} + g_{qr}\frac{\partial x^q}{\partial x'^k}\frac{\partial^2 x^r}{\partial x'^j x'^l} + g_{qr}\frac{\partial x^r}{\partial x'^j}\frac{\partial^2 x^q}{\partial x'^k x'^l} + g_{qr}\frac{\partial x^q}{\partial x'^l}\frac{\partial^2 x^r}{\partial x'^k x'^j} - g_{qr}\frac{\partial x^r}{\partial x'^k}\frac{\partial^2 x^q}{\partial x'^l x'^j} - g_{qr}\frac{\partial x^q}{\partial x'^j}\frac{\partial^2 x^r}{\partial x'^l x'^k}\right) \tag{136}$$

$$= \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}\Gamma^l{}_{mn} + \frac{\partial x'^i}{\partial x^m}\frac{\partial x'^l}{\partial x^n} g^{mn} g_{qr}\frac{\partial x^r}{\partial x'^l}\frac{\partial^2 x^q}{\partial x'^j x'^k} \tag{137}$$

$$= \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}\Gamma^l{}_{mn} + \frac{\partial x'^i}{\partial x^m} g^{mn} g_{qr}\delta^r{}_n\frac{\partial^2 x^q}{\partial x'^j x'^k} \tag{138}$$

$$= \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}\Gamma^l{}_{mn} + \frac{\partial x'^i}{\partial x^m}\frac{\partial^2 x^m}{\partial x'^j x'^k} \tag{139}$$

which matches theorem 2.4. □

The substance of the previous definition is sometimes also given the infinitely grander title of the "fundamental theorem of Riemannian geometry."

Definition 2.15 (Riemann tensor). *The local curvature 2-form of a connection on a manifold (not necessarily Levi-Civita) is called the Riemann tensor.*

Theorem 2.16. *The Riemann tensor really is a tensor across the whole manifold and is given in local coordinates by*

$$R^i{}_{jkl} = \partial_k(\Gamma^i{}_{jl}) - \partial_l(\Gamma^i{}_{jk}) + \Gamma^i{}_{mk}\Gamma^m{}_{jl} - \Gamma^i{}_{ml}\Gamma^m{}_{jk} \tag{140}$$

where k and l are the 2-form indices and i and j are the $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$ indices.

Proof. The local curvature 2-form is defined in local coordinates be equation 79. Restoring the suppressed Lie algebra indices (just matrix indices in the case of $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$) in that equation, letting Γ_k denote the matrix whose components are $\Gamma^i{}_{jk}$ and noting that the Lie bracket is just matrix commutator for $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$, equation 79 says

$$R^i{}_{jkl} = \partial_k(\Gamma^i{}_{jl}) - \partial_l(\Gamma^i{}_{jk}) + [\Gamma_k, \Gamma_l]^i{}_j \tag{141}$$

$$= \partial_k(\Gamma^i{}_{jl}) - \partial_l(\Gamma^i{}_{jk}) + \Gamma^i{}_{mk}\Gamma^m{}_{jl} - \Gamma^i{}_{ml}\Gamma^m{}_{jk} \tag{142}$$

which proves the second half of the theorem. For the first half, applying theorem 1.27 with $g_{\beta\alpha} = \partial x'/\partial x$ implies that upon a change of coordinates,

$$R'^i{}_{jkl}\mathrm{d}x'^k\mathrm{d}x'^l = \frac{\partial x'^i}{\partial x^m}R^m{}_{nkl}\frac{\partial x^n}{\partial x'^j}\mathrm{d}x^k\mathrm{d}x^l \tag{143}$$

$$= \frac{\partial x'^i}{\partial x^m}\frac{\partial x^n}{\partial x'^j}\frac{\partial x^p}{\partial x'^k}\frac{\partial x^q}{\partial x'^l}R^m{}_{npq}\mathrm{d}x'^k\mathrm{d}x'^l \tag{144}$$

$$\therefore R'^i{}_{jkl} = \frac{\partial x'^i}{\partial x^m}\frac{\partial x^n}{\partial x'^j}\frac{\partial x^p}{\partial x'^k}\frac{\partial x^q}{\partial x'^l}R^m{}_{npq} \tag{145}$$

which is exactly the required transformation property on overlaps to ensure that $R^i{}_{jkl}$ forms a type-(1, 3) tensor globally on the manifold. □

Theorem 2.17. *For a Levi-Civita connection, the Riemann tensor vanishes if and only if* $\exists$ *local coordinates,* x^i*, such that* $g = \eta_{ij}\mathrm{d}x^i \otimes \mathrm{d}x^j$*.*

Proof. First, assume $g = \eta_{ij}\mathrm{d}x^i \otimes \mathrm{d}x^j$.
Then, since $g_{ij} = \eta_{ij}$ are constants, by equation 133, $\Gamma^i{}_{jk} = 0$ and thus by theorem 2.16, $R^i{}_{jkl} = 0$.
For the much harder direction, first assume $R = 0$.
On principle bundles, as proven earlier, local sections are equivalent to trivialisations.
$\therefore F_\alpha = 0 \iff \mathcal{F} = 0$.
In this case, that means $R = 0$ if and only if the corresponding curvature on the principal bundle is zero.
Then, by theorems 1.25 and 1.22, $R = 0 \implies \Gamma_\alpha = 0$ in some local basis, $\{e_i\}_{i=1}^n$, for the tangent bundle[8]. Let $\{\varepsilon^i\}_{i=1}^n$ be the dual basis over this trivialising patch.
In any basis, by definition, $\nabla(v^i e_i) = \mathrm{d}(v^i)e_i + (\Gamma_\alpha)^i{}_j v^j e_i$, where I'm only showing the $\mathfrak{gl}(\mathfrak{n}, \mathbb{R})$ indices.
$\therefore \nabla(e_i) = (\Gamma_\alpha)^j{}_i e_j$, which in this case is just zero because $\Gamma_\alpha = 0$.
Then, applying corollary 2.7.2 to the present torsion-free connection,
$[e_i, e_j] = \nabla_{e_i} e_j - \nabla_{e_j} e_i = 0 - 0 = 0$.
A local tangent bundle basis is known (e.g. from Frobenius' theorem) to be commuting if and only if it is coordinate induced, so it follows that $e_i = \partial_i$ for some local coordinates, x^i. Now I can apply the coordinate expressions for the Christoffel symbols deduced in equation 133.

$$\begin{aligned} \therefore 0 &= \Gamma_{ijk} + \Gamma_{jik} && (146)\\ &= \frac{1}{2}(\partial_j g_{ki} + \partial_k g_{ij} - \partial_i g_{jk} + \partial_i g_{kj} + \partial_k g_{ji} - \partial_j g_{ik}) && (147)\\ &= \partial_k g_{ij} && (148) \end{aligned}$$

Since the metric components are locally constant, I can perform a linear change of coordinates, $x'^i = M^i{}_j x^j$, such that

$$g'_{ij} = \frac{\partial x^k}{\partial x'^i}\frac{\partial x^l}{\partial x'^j} g_{kl} = (M^{-1})^k{}_i (M^{-1})^l{}_j g_{kl} = \eta_{ij} \tag{149}$$

throughout the coordinate patch. $\square$

Theorem 2.18. *For an arbitrary connection, the covariant derivatives satisfy the commutatator identity,*

$$\begin{aligned} [\nabla_i, \nabla_j] A^{k_1 \cdots k_p}{}_{l_1 \cdots l_q} = -T^m{}_{ij}\nabla_m A^{k_1 \cdots k_p}{}_{l_1 \cdots l_q} + \sum_{a=1}^{p} R^{k_a}{}_{mij} A^{k_1 \cdots \hat{k_a} m \cdots k_p}{}_{l_1 \cdots l_q} \\ - \sum_{a=1}^{q} R^m{}_{l_a ij} A^{k_1 \cdots k_p}{}_{l_1 \cdots \hat{l_a} m l_q} \end{aligned} \tag{150}$$

where A *is an arbitrary tensor field and* $T^m{}_{ij}$ *are the components of the torsion tensor.*

Proof. Both sides are local, so it suffices to check the statement in local coordinates. I'll start

[8] i.e. the $\{e_i\}_{i=1}^n$ is a fibrewise basis generated from the collection of local sections, $\{s_i\}_{i=1}^n$, that trivialises TM over the patch, V_α, in which $\Gamma_\alpha = 0$.

by only considering A to be a scalar, vector or covector. Using theorems 2.5, 2.7 and 2.16,

$$
\begin{aligned}
&[\nabla_i, \nabla_j]A \\
&= \nabla_i\nabla_j A - \nabla_j\nabla_i A && (151)\\
&= \partial_i(\nabla_j A) - \Gamma^k{}_{ji}\nabla_k A - \partial_j(\nabla_i A) - \Gamma^k{}_{ij}\nabla_k A && (152)\\
&= \partial_i\partial_j A - \Gamma^k{}_{ji}\partial_k A - \partial_j\partial_i A - \Gamma^k{}_{ij}\partial_k A && (153)\\
&= (\Gamma^k{}_{ij} - \Gamma^k{}_{ji})\partial_k A && (154)\\
&= -T^k{}_{ij}\nabla_k A && (155)\\
&[\nabla_i\nabla_j]A^k \\
&= \nabla_i\nabla_j A^k - \nabla_j\nabla_i A^k && (156)\\
&= \partial_i(\nabla_j A^k) - \Gamma^m{}_{ji}\nabla_m A^k + \Gamma^k{}_{mi}\nabla_j A^m - \partial_j(\nabla_i A^k) + \Gamma^m{}_{ij}\nabla_m A^k - \Gamma^k{}_{mj}\nabla_i A^m && (157)\\
&= -T^m{}_{ij}\nabla_m A^k + \partial_i\partial_j A^k + \partial_i(\Gamma^k{}_{mj})A^m + \Gamma^k{}_{mj}\partial_i A^k + \Gamma^k{}_{mi}\partial_j A^k + \Gamma^k{}_{mi}\Gamma^m{}_{nj}A^n \\
&\quad - \partial_j\partial_i A^k - \partial_j(\Gamma^k{}_{mi})A^m - \Gamma^k{}_{mi}\partial_j A^k - \Gamma^k{}_{mj}\partial_i A^k - \Gamma^k{}_{mj}\Gamma^m{}_{ni}A^n && (158)\\
&= -T^m{}_{ij}\nabla_m A^k + (\partial_i(\Gamma^k{}_{mj}) + \Gamma^k{}_{ni}\Gamma^n{}_{mj} - \partial_j(\Gamma^k{}_{mi}) - \Gamma^k{}_{nj}\Gamma^n{}_{mi})A^m && (159)\\
&= -T^m{}_{ij}\nabla_m A^k + R^k{}_{mij}A^m && (160)\\
&[\nabla_i, \nabla_j]A_k \\
&= \nabla_i\nabla_j A_k - \nabla_j\nabla_i A_k && (161)\\
&= \partial_i(\nabla_j A_k) - \Gamma^m{}_{ji}\nabla_m A_k - \Gamma^m{}_{ki}\nabla_j A_m - \partial_i(\nabla_j A_k) + \Gamma^m{}_{ji}\nabla_m A_k + \Gamma^m{}_{ki}\nabla_j A_m && (162)\\
&= -T^m{}_{ij}\nabla_m A_k + \partial_i\partial_j A_k - \partial_i(\Gamma^m{}_{kj})A_m - \Gamma^m{}_{kj}\partial_i A_m - \Gamma^m{}_{ki}\partial_j A_m + \Gamma^m{}_{ki}\Gamma^n{}_{mj}A_n \\
&\quad - \partial_j\partial_i A_k + \partial_j(\Gamma^m{}_{ki})A_m + \Gamma^m{}_{ki}\partial_j A_m + \Gamma^m{}_{kj}\partial_i A_m - \Gamma^m{}_{kj}\Gamma^n{}_{mi}A_n && (163)\\
&= -T^m{}_{ij}\nabla_m A_k - (\partial_i(\Gamma^m{}_{kj}) - \Gamma^n{}_{ki}\Gamma^m{}_{nj} - \partial_j(\Gamma^m{}_{ki}) + \Gamma^n{}_{kj}\Gamma^m{}_{ni})A_m && (164)\\
&= -T^m{}_{ij}\nabla_m A_k - R^m{}_{kij}A_m && (165)
\end{aligned}
$$

which confirms the theorem in these three special cases.
However, because of the way the covariant derivative acts "index by index" in theorem 2.5, the general result follows immediately. □

The Riemann tensor possesses many symmetries that are not manifestly apparent. There are often deep reasons for these symmetries, but I will not bother cataloguing those reasons here.

Theorem 2.19. *The Riemann tensor satisfies the following properties.*

1. $R^i{}_{jkl} = -R^i{}_{jlk}$ *for an arbitrary connection*
2. $R_{ijkl} = -R_{jikl}$ *for the Levi-Civita connection*
3. $R_{ijkl} = R_{klij}$ *for the Levi-Civita connection*
4. $R^i{}_{[jkl]} = 0$ *for a torsion-free connection (1st Bianchi identity)*
5. $\nabla_{[i}R^j{}_{|k|lm]} = 0$ *for a torsion-free connection (2nd Bianchi identity)*

Proof. (1) is immediate because k and l are the two-form indices of the curvature and are thus automatically antisymmetric.
Next, assume the connection is torsion-free. Then, from theorem 2.18, for an arbitrary vector field, V,

$$
\begin{aligned}
[\nabla_i, [\nabla_j, \nabla_k]]V^l &= \nabla_i(R^l{}_{mjk}V^m) - [\nabla_j, \nabla_k]\nabla_i V^l && (166)\\
&= \nabla_i(R^l{}_{mjk})V^m + R^l{}_{mjk}\nabla_i V^m - R^l{}_{mjk}\nabla_i V^m + R^m{}_{ijk}\nabla_m V^l && (167)\\
&= \nabla_i(R^l{}_{mjk})V^m + R^m{}_{ijk}\nabla_m V^l && (168)
\end{aligned}
$$

Since commutators satisfy the Jacobi identity, it follows that

$$
\begin{aligned}
0 &= [\nabla_i, [\nabla_j, \nabla_k]]V^l + [\nabla_j, [\nabla_k, \nabla_i]]V^l + [\nabla_k, [\nabla_i, \nabla_j]]V^l && (169)\\
&= \nabla_i(R^l{}_{mjk})V^m + R^m{}_{ijk}\nabla_m V^l + \nabla_j(R^l{}_{mki})V^m + R^m{}_{jki}\nabla_m V^l \\
&\quad + \nabla_k(R^l{}_{mij})V^m + R^m{}_{kij}\nabla_m V^l && (170)\\
&= (\nabla_i(R^l{}_{mjk}) + \nabla_j(R^l{}_{mki}) + \nabla_k(R^l{}_{mij}))V^m + (R^m{}_{ijk} + R^m{}_{jki} + R^m{}_{kij})\nabla_m V^l && (171)
\end{aligned}
$$

Then, because V is arbitrary, properties (4) and (5) must be true.
Finally, let the connection be Levi-Civita.
Let V and W be arbitrary tensor fields. Then, $g_{ij}V^iW^j$ is a scalar and theorem 2.18 implies

$$
\begin{aligned}
0 &= [\nabla_i, \nabla_j](g_{kl}V^kW^l) && (172)\\
&= g_{kl}[\nabla_i, \nabla_j](V^kW^l) \text{ as } \nabla_i g_{jk} = 0 \text{ for the Levi} - \text{Civita connection} && (173)\\
&= g_{kl}(\nabla_i(\nabla_j(V^k)W^l + V^k\nabla_j W^l) - \nabla_j(\nabla_i(V^k)W^l + V^k\nabla_i W^l)) && (174)\\
&= g_{kl}(\nabla_i\nabla_j(V^k)W^l + \nabla_j(V^k)\nabla_i(W^l) + \nabla_i(V^k)\nabla_j(W^l) + V^k\nabla_i\nabla_j W^l \\
&\quad - \nabla_j\nabla_i(V^k)W^l + \nabla_i(V^k)\nabla_j(W^l) + \nabla_j(V^k)\nabla_i(W^l) + V^k\nabla_j\nabla_i W^l) && (175)\\
&= g_{kl}(W^l[\nabla_i, \nabla_j]V^k + V^k[\nabla_i, \nabla_j]W^l) && (176)\\
&= g_{kl}(R^k{}_{mij}V^mW^l + R^l{}_{mij}W^mV^k) && (177)\\
&= (R_{lkij} + R_{klij})V^kW^l && (178)
\end{aligned}
$$

$\therefore$ As V and W are arbitrary, property (2) must hold.
Then, liberally utilising properties (1), (2) and (4),

$$
\begin{aligned}
R_{ijkl} &= -R_{iklj} - R_{iljk} && (179)\\
&= R_{kilj} + R_{lijk} && (180)\\
&= -R_{klji} - R_{kjil} - R_{ljki} - R_{lkij} && (181)\\
&= 2R_{klij} + R_{jkil} + R_{jlki} && (182)\\
&= 2R_{klij} - R_{jilk} && (183)\\
&= 2R_{klij} - R_{ijkl} && (184)\\
\therefore R_{ijkl} &= R_{klij} && (185)
\end{aligned}
$$

which is property (3). □

Definition 2.20 (Ricci tensor and scalar)**.** *The Ricci tensor and Ricci scalar are defined by* $R_{ij} = R^k{}_{ikj}$ *and* $R = R^i{}_i$ *respectively.*

Theorem 2.21. *When using a Levi Civita connection, the Ricci tensor and scalar satisfy*

1. $R_{ij} = R_{ji}$
2. $\nabla^j R_{ij} = \frac{1}{2}\nabla_i R$

Proof. Using the Riemann tensor's algebraic symmetries, $R_{ij} = R^k{}_{ikj} = R_{kj}{}^k{}_i = R^k{}_{jki} = R_{ji}$, which is property (1).
For property (2), contracting indices on the 2nd Bianchi identity and then applying the algebraic symmetries and using property (1),
$0 = \nabla^i R^{jk}{}_{jk} + \nabla_j R^{jk}{}_k{}^i + \nabla_k R^{jki}{}_j = \nabla^i R - \nabla_j R^{ji} - \nabla_k R^{ki} \implies \nabla^j R_{ij} = \frac{1}{2}\nabla_i R.$ □

2.2 Non-coordinate bases

2.3 Yang-Mills theory

2.4 Conformal geometry