4.7 The Levi-Civita connection and parallel transport

In the earlier investigation, characterizing the shortest curves between two points was cast as a variational problem. The resulting necessary condition has the form of a system of second order differential equations. As a consequence, geodesics, as solutions of smooth initial value problems automatically have substantial regularity properties.

This section takes a closer look at the coefficients that appeared in the Euler Lagrange equation of the variational problem: These are linear combinations of partial derivatives of the Riemannian metric. Intuitively they measure how much the metric varies from point to point. Or even more markedly, how much the difference between the metric on the manifold from the flat Euclidean manifold varies from point to point. Clearly such rates of change of the Riemannian metric are predestined to lead to an intrinsic curvature of the manifold.

The plan for the upcoming discussions is to revisit the Christoffel symbols as derivatives of the metric ($\Gamma^{i}_{jk}$), and to attribute rates of change of the metric to rates of change of vector fields. As a first step, we explore the rates of change (directional derivatives) of coordinate vector fields. This development shall be guided by a close inspection of derivatives in/of curvilinear coordinates in Euclidean spaces, and exploration of derivatives of tangent vector fields on hypersurfaces. Eventually this leads to an axiomatic definition of covariant derivatives or connections (or connexions). The last part of this section is devoted to the discussion of some technical properties of this derivative. Subsequent sections shall take this notion of connection to develop elegant characterizations of curvature.

The previous section introduced the Christoffel symbols in equation (216)

$$[i,j,k] = \frac{1}{2} \left( \frac{\partial g_{ik}}{\partial u^j} + \frac{\partial g_{jk}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k} \right)$$

basically as abbreviations for certain linear combinations of partial derivatives of the Riemannian metric. Using the symmetry of the Riemannian metric it is straightforward to solve backwards to express the partial derivatives of the Riemannian metric as linear combinations of the Christoffel symbols:

$$[i,j,k] + [jk,i] = \frac{1}{2} \left( \left( \frac{\partial g_{ik}}{\partial u^j} + \frac{\partial g_{jk}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k} \right) + \left( \frac{\partial g_{ji}}{\partial u^k} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{jk}}{\partial u^i} \right) \right) = \frac{\partial g_{ik}}{\partial u^j}$$

What matters to us is that the coefficients in the geodesic equation are indeed, essentially, the rates of change of the Riemannian metric: If (in a suitable coordinate chart) the metric is constant, then $\frac{d^2 (u \circ \sigma)}{dt^2} \equiv \text{const}$. This says that if a manifold is flat, then the shortest curves are (straight) lines.

Recall that the components $g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle$ of the metric are the inner products of the coordinate vector fields. One goal of this section is to attribute the rate of change of the Riemannian metric, in a geometrically consistent way, to the rates of change of the basic vector fields themselves. Given the characteristic features of any notion of derivative, linearity and being a derivation (i.e., satisfying Leibniz’ rule or the product rule), it is not far-fetched to ask that such a notion of a derivative $\nabla_Z$ (in the direction of a vector field $Z$) satisfy (compare theorem (287))

$$Z g_{ij} = Z \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle = \langle \nabla_Z \left( \frac{\partial}{\partial u^i} \right), \frac{\partial}{\partial u^j} \rangle + \langle \frac{\partial}{\partial u^i}, \nabla_Z \left( \frac{\partial}{\partial u^j} \right) \rangle$$

In the end, this compatibility condition connecting notions of rates of change of vector fields to the rate of change of the metric plays a critical role in singling out a special covariant derivative among many possible others.
This leads to the question about what one might mean by the rate of change of a vector field. In some sense, this corresponds to, instead of focusing on shortest, to study notions of straight and parallel. Before reading on, it is recommended to spend some effort on the following exploration:

**Exercise 4.65** Explore/develop a notion of the rate of change of the coordinate vector fields \( \frac{\partial}{\partial r} = \cos \theta D_1 + \sin \theta D_2 \) and \( \frac{\partial}{\partial \theta} = -r \sin \theta D_1 + r \cos \theta D_2 \) with respect to (“in the directions of”) each other, and in the directions of \( D_1 \) and \( D_2 \). [[This is much simpler than it looks – very simple and intuitive ideas for directional derivatives are the right thing to do!]].

One basic observation is that the standard coordinate vector fields \( D_i \) in Euclidean space serve as models for straightness, for constancy, or parallelism. Consequently, for basically any notion of derivative, the derivative of these fields ought to be zero (although not for Lie derivatives, see below).

In terms of the preceding exercise, compare how \( \frac{\partial}{\partial \theta} \) changes in the direction of \( D_1 = \frac{\partial}{\partial x} \) with how \( D_1 \) changes in the direction \( \frac{\partial}{\partial \theta} \). However, note that the Lie derivative satisfies \( L_{\frac{\partial}{\partial \theta}} D_1 = -L_{\frac{\partial}{\partial x}} \frac{\partial}{\partial \theta} \), and thus does not have the desired asymmetry that reflects that \( D_1 \) as straight while \( \frac{\partial}{\partial \theta} \) as not straight. In the Euclidean context it is natural to define

**Definition 4.16** The Euclidean connection (or Euclidean covariant derivative) is the map 
\[ \nabla : \Gamma^\infty(\mathbb{R}^m) \times \Gamma^\infty(\mathbb{R}^m) \mapsto \Gamma^\infty(\mathbb{R}^m) \] defined by \( \nabla_Z \left( \sum_{i=1}^{m} X^i D_i \right) = \sum_{i=1}^{m} Z(X^i)D_i \).

Looking ahead to general manifolds where there is no reason to expect that a typical (“any”) vector field be constant in any reasonable sense, the defining equation may be written in a longer, more suggestive form as:

\[ \nabla_Z(X) = \nabla_Z \left( \sum_{i=1}^{m} X^i D_i \right) = \sum_{i=1}^{m} Z(X^i)D_i + X^i(\nabla_Z(D_i)) = \sum_{i=1}^{m} Z(X^i)D_i \quad (289) \]

The next objective is to develop similar notions for derivatives on manifolds. These should be intrinsic, i.e. not utilize an imbedding into an ambient Euclidean space. Nonetheless, coordinate changes and surfaces in \( \mathbb{R}^3 \) provide valuable guidance on how to define covariant derivatives on general manifolds.

**Exercise 4.66** Let \((x^1, x^2)\) denote the standard coordinates on \( \mathbb{R}^2 \) and suppose \((u^1, u^2) = (\Phi^1 \circ (x^1, x^2), \Phi^2 \circ (x^1, x^2))\) is another set of smooth coordinates. Expand the coordinate vector fields \( \frac{\partial}{\partial u^i} \) as linear combinations of \( D_j \) with partial derivatives of \( \Phi^k \) as coefficients. Calculate the matrix \( g_{ij} \) representing the standard Riemannian, i.e. Euclidean, metric w.r.t. the coordinates \((u^1, u^2)\). Use the identities \( \nabla_{D_i} D_j \equiv 0 \) and equation (??) to calculate \( \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} \).
Next explore vector fields on hypersurfaces in Euclidean spaces. Working locally, assume w.l.o.g. that the manifold is the graph of a function \( x^{m+1} = f(x^1, x^2, \ldots, x^m) \). For the sake of clarity consider the special case \( m = 2 \) (note that if \( m = 2 \) there necessarily is some repetition among the indices of \( \Gamma^k_{ij} \)). Suppose \( M^2 \subseteq \mathbb{R}^3 \) is the graph of a smooth function \( x^3 = f(x^1, x^2) \). Using the local coordinates \( u(x) = (x^1, x^2) \) the coordinate vector fields and induced metric are

\[
\frac{\partial}{\partial u^1} = D_1 + (D_1 f) D_3 \quad \text{and} \quad (g_{ij})_{i,j=1,2} = \begin{pmatrix} 1 + (D_1 f)^2 & (D_1 f) \cdot (D_2 f) \\ (D_1 f) \cdot (D_2 f) & 1 + (D_2 f)^2 \end{pmatrix}
\]

(290)

Using the Euclidean connection (??), and identifying the vector fields \( \frac{\partial}{\partial u^i} \) on \( M \) with the natural extensions of the vector fields \( \iota_* \frac{\partial}{\partial u^i} \) on \( \iota_* M \) to vector fields on \( \mathbb{R}^3 \), calculate

\[
\nabla \frac{\partial}{\partial u^1} = \nabla_{D_1 + (D_1 f) D_3} (D_1 + (D_1 f) D_3) = (D_1^2 f) D_3
\]

\[
\nabla \frac{\partial}{\partial u^2} = \nabla_{D_2 + (D_2 f) D_3} (D_1 + (D_1 f) D_3) = (D_2 D_1 f) D_3
\]

\[
\nabla \frac{\partial}{\partial u^3} = \nabla_{D_1 + (D_1 f) D_3} (D_2 + (D_2 f) D_3) = (D_1 D_2 f) D_3
\]

\[
\nabla \frac{\partial}{\partial u^3} = \nabla_{D_2 + (D_2 f) D_3} (D_2 + (D_2 f) D_3) = (D_2^2 f) D_3.
\]

(291)

In general there is no reason for these vector fields on \( \mathbb{R}^{m+1} \) to restrict to vector fields on \( M^m \). However, in the case of smooth hypersurfaces \( M^m \subseteq \mathbb{R}^{m+1} \) it is natural to extend the set of coordinate vector fields on \( M^m \) to a frame on \( \mathbb{R}^{m+1} \) by adjoining a normal vector field \( \nu \). The existence of such normal vector field depends crucially on the Riemannian metric on the ambient space. In the case of \( m = 2 \) this yields

\[
\frac{\partial}{\partial u^1} = D_1 + (D_1 f) D_3 \\
\frac{\partial}{\partial u^2} = D_2 + (D_2 f) D_3 \\
\nu = - (D_1 f) D_1 - (D_2 f) D_2 + D_3.
\]

(292)

It is straightforward to invert this system, to solve for \( D_i \) in terms of \( \frac{\partial}{\partial u^i} \) and \( \nu \)

\[
D_1 = \frac{1 + (D_1 f)^2}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} - \frac{(D_1 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} - \frac{(D_1 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu
\]

\[
D_2 = - \frac{1 + (D_2 f)^2}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} + \frac{(D_2 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} - \frac{(D_2 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu
\]

\[
D_3 = \frac{(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} + \frac{(D_2 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} - \frac{1}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu.
\]

(293)

Use these to rewrite the derivatives (??) as linear combinations of \( \frac{\partial}{\partial u^i} \) and \( \nu \)

\[
\nabla \frac{\partial}{\partial u^1} = \frac{(D_2 f)(D_1 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} + \frac{(D_2 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} + \frac{(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu
\]

\[
\nabla \frac{\partial}{\partial u^2} = \frac{(D_1 f)(D_2 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} + \frac{(D_2 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} + \frac{(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu
\]

\[
\nabla \frac{\partial}{\partial u^3} = \frac{(D_1 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^1} + \frac{(D_2 f)(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \frac{\partial}{\partial u^2} + \frac{(D_3 f)}{1 + (D_1 f)^2 + (D_2 f)^2} \cdot \nu.
\]

(294)
One recognizes the coefficients of the vector fields $\frac{\partial}{\partial u^i}$ in these expansions as the Christoffel symbols of the second kind, i.e. for suitable functions $c_{ij}$

$$\nabla \frac{\partial}{\partial u^i} \frac{\partial}{\partial u^j} = \sum_k \Gamma^k_{ij} \frac{\partial}{\partial u^k} + c_{ij} \nu$$  \hspace{1cm} (295)

This observation suggests to define an induced connection on hypersurfaces by

$$\nabla \frac{\partial}{\partial u^i} \frac{\partial}{\partial u^j} = \sum_k \Gamma^k_{ij} \frac{\partial}{\partial u^k}.$$  \hspace{1cm} (296)

Of course it needs to be checked that this is indeed well-defined, i.e. independent of the coordinate chart. We will postpone this until after the definition of connections on any smooth manifold, not only on hypersurfaces.

**Exercise 4.67** Verify that with this definition the following identity holds:

$$\frac{\partial}{\partial u^i} (g_{j\ell}) = \langle \nabla \frac{\partial}{\partial u^i} \left( \frac{\partial}{\partial u^j} \right), \frac{\partial}{\partial u^\ell} \rangle + \langle \frac{\partial}{\partial u^j}, \nabla \frac{\partial}{\partial u^i} \left( \frac{\partial}{\partial u^\ell} \right) \rangle$$  \hspace{1cm} (297)

These formulas also lead to an intuitive interpretation of the role of the Christoffel symbols in the geodesic equations: In the case of hypersurfaces $M^m \subseteq R^{m+1}$ one may consider geodesics $\sigma: [a, b] \mapsto M^m$ as curves in $R^{m+1}$. In general their acceleration $\sigma''$ will not be zero – i.e. they will not be straight lines (traversed at constant speed) in $R^{m+1}$ – but the acceleration $\sigma''(t)$ is always perpendicular to the tangent plane $T_\sigma(t)M^m$.

**Exercise 4.68** Prove this assertion, i.e. suppose that $\sigma: [a, b] \mapsto M^m \subseteq R^{m+1}$ is a geodesic on a hypersurface. Show that the “acceleration” $(\nu \circ \sigma)''(t)$, considered as a curve in $T_{\sigma(t)}R^{m+1}$, is always orthogonal to $T_{\sigma(t)}M^m \subseteq T_{\sigma(t)}R^{m+1}$. [Compare also the MAPLE worksheets.]

It is time for a formal definition – the following invariant definition is usually associated with the name Koszul connection. Note, that this definition does not require a metric – but at this time we shall only consider it in the setting of Riemannian manifolds.

**Definition 4.17** Suppose $M$ is a smooth manifold. An (affine) connection (or “covariant derivative” on $M$) is a map $\nabla: \Gamma^\infty(M) \times \Gamma^\infty(M) \mapsto \Gamma^\infty(M)$ that satisfies for all $X, Y, Z \in \Gamma^\infty(M)$, all $f \in C^\infty(M)$ and all $c \in \mathbb{R}$

(i) $\nabla_{fY + Z}X = f\nabla_YX + \nabla_ZX$,

(ii) $\nabla_Z(cX + Y) = c\nabla_ZX + \nabla_ZY$,

(iii) $\nabla_Z(fX) = (Zf)X + f\nabla_ZX$.

**Definition 4.18** An affine connection $\nabla$ on a Riemannian manifold $M$ is called a Riemannian connection if it in addition to (i)-(iii), it also satisfies, for all $X, Y, Z \in \Gamma^\infty(M)$,

(iv) $\nabla_X Y - \nabla_Y X = [X, Y]$, and

(v) $Z\langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle$.

We note on the side that these definitions immediately generalize to covariant derivatives of any tensor. Moreover, covariant derivatives commute with contractions of tensors (for details on both compare Spivak II, 2nd ed., p.244).
Theorem 4.42 A Riemannian manifold admits exactly one Riemannian connection.

Proof. This is an immediate consequence from the following calculation which expresses the covariant derivative entirely in terms of the Riemannian metric. For vector fields $X, Y, Z \in \Gamma^\infty(M)$ use property (v):

$$
X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle
$$

Suitably adding and subtracting the right and left hand sides and using property (iv) yields

$$
\langle \nabla_Y Z, X \rangle = \frac{1}{2} \left( \langle [X, Y], Z \rangle + \langle [X, Z], Y \rangle + \langle [Y, Z], X \rangle - X \langle Y, Z \rangle + Y \langle Z, X \rangle + Z \langle X, Y \rangle \right)
$$

Since the inner product is nondegenerate this completely determines $\nabla_Y Z$.

Conversely on a Riemannian manifold $(R, \langle \cdot, \cdot \rangle)$ define in any chart $(u, U)$ with Christoffel symbols $\Gamma^k_{ij}$ as defined in equation (216)

$$
\nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} = \sum_{k=1}^m \Gamma^k_{ij} \frac{\partial}{\partial u^k}
$$

and extend to all vector fields by demanding the properties (i) through (iii). Due to the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ the left hand side of (iv) vanishes in the case of coordinate vector fields $X = \frac{\partial}{\partial u^i}$ and $Y = \frac{\partial}{\partial u^j}$, as does the right hand side. Finally, (v) follows from

$$
\langle \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j}, \frac{\partial}{\partial u^k} \rangle + \langle \frac{\partial}{\partial u^j}, \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^k} \rangle = \sum_{\ell=1}^m \Gamma^\ell_{ki} \cdot g_{\ell j} + \Gamma^\ell_{kj} \cdot g_{\ell i}
$$

Exercise 4.69 Establish the classical transformation formula for the Christoffel symbols of the second kind: Suppose that $(u, U)$ and $(v, V)$ are charts with metrics $g_{ij}$ and $\tilde{g}_{\alpha\beta}$, and Christoffel symbols $\Gamma^k_{ij}$ and $\tilde{\Gamma}^\gamma_{\alpha\beta}$ respectively. Then:

$$
\tilde{\Gamma}^\gamma_{\alpha\beta} = \sum_{i,j,k=1}^m \Gamma^k_{ij} \cdot \frac{\partial u^i}{\partial v^\alpha} \frac{\partial u^j}{\partial v^\beta} \frac{\partial v^\gamma}{\partial u^k} + \sum_{s=1}^m \frac{\partial^2 u^s}{\partial v^\alpha \partial v^\beta} \frac{\partial v^\gamma}{\partial u^k}.
$$
In a chart \((u, U)\) calculate for general vector fields \(X = \sum_{i=1}^{m} (Xu^i) \frac{\partial}{\partial u^i}\) and \(Y = \sum_{j=1}^{m} (Yu^j) \frac{\partial}{\partial u^j}\)

\[
\nabla_X Y = \nabla_X \left( \sum_{j=1}^{m} (Yu^j) \frac{\partial}{\partial u^j} \right)
= \sum_{j=1}^{m} \left( (XYu^j) \frac{\partial}{\partial u^j} + \sum_{i=1}^{m} (Xu^i) (Yu^j) \nabla_\frac{\partial}{\partial u^i} \frac{\partial}{\partial u^j} \right)
= \sum_{k=1}^{m} \left( (XYu^k) + \sum_{i=1}^{m} \Gamma_{ij}^k (Xu^i) (Yu^j) \right) \frac{\partial}{\partial u^k}
\]

(304)

This expression looks remarkably similar to the geodesic equations – but there are some technical issues that need to be addressed: If \(\sigma: [a, b] \mapsto M\) is a smooth curve then \(\sigma' = \sigma \frac{d}{dt}\) is defined only along the curve, i.e. it is not immediately clear in which sense one may take \(X = Y = \sigma'\) in above formula. The following exercise suggests that one can indeed make sense of a derivative of a vector field defined only along a curve.

**Exercise 4.70** Suppose \(p \in M\) and \(X, Y \in \Gamma^\infty(p)\) are smooth vector fields defined in a neighbourhood of \(p\). Show that \((\nabla_X Y)_p\) is completely determined by \(X_p\) and the values of \(Y\) along any curve \(\sigma: (-\varepsilon, \varepsilon) \mapsto M\) such that \(\sigma'(0) = X_p\). [[Suppose that \(\sigma\) is a curve as described above, and \(X', Y'\) are vector fields such that \(X'_p = X_p\) and for all \(t \in (-\varepsilon, \varepsilon), Y'_{\sigma(t)} = Y_{\sigma(t)}\). Use these to show that \((\nabla_X Y)_p = (\nabla_X Y')_p\) and that \((\nabla_X Y')_p = (\nabla_X Y)_p\). Analyze the differences and use the axioms (??), e.g. (iii) with \(f\) such \(X_p f = 0\)]]

**Definition 4.19** Suppose \(\sigma: [a, b] \mapsto M\) is a curve. A vector field along the curve \(\sigma\) is a function \(X: [a, b] \mapsto TM\) such that \(\pi \circ X = \sigma\).

It is critical that \(X\) is defined on \([a, b]\) as opposed to being defined on the subset \(\sigma([a, b]) \subseteq M\). In particular, if the curve is not one-to-one, e.g. if \(\sigma(t_1) = \sigma(t_2)\) for some \(t_1 \neq t_2\) it is permissible to have \(X_{t_1} \neq X_{t_2}\). In general our interest lies in smooth curves, and smooth vector fields along a curve – defined as smooth maps between manifolds \([a, b] \subseteq \mathbb{R}\) and \(TM\).

In analogy to vector fields along a curve one may define vector fields along a hypersurfaces and similar objects – but we shall not have any immediate need for these. [[Some discussions of curvature make use of vector fields defined on rectangles \(X: (-\varepsilon, \varepsilon) \times [0, 1] \mapsto M\) in conjunction with maps \(H: (-\varepsilon, \varepsilon) \times [0, 1] \mapsto M\), i.e., homotopies between curves \(H(\cdot, 0)\) and \(H(\cdot, 1)\).]]

**Exercise 4.71** Suppose that \(X: [a, b] \mapsto TM\) is a smooth vector field along a smooth curve \(\sigma: [a, b] \mapsto M\) such that for all \(t \in [a, b]\), \(\sigma'(t) \neq 0\). Show that for any \(t \in [a, b]\) there exists \(\varepsilon > 0\), an open neighborhood \(U\) of \(\sigma([-t + \varepsilon, t + \varepsilon]) \subseteq M\) and a vector field \(\overline{X}: U \mapsto TM\) such that \(\overline{X} \circ \sigma = X\) on \([t - \varepsilon, t + \varepsilon]\). [[Use the immersion theorem (2.13).]]
Proposition 4.43 For every smooth curve $\sigma : [a, b] \mapsto M$ there exists exactly one operation, denoted $\frac{D}{dt}$, from $C^\infty$-vector fields along $\sigma$ to $C^\infty$-vector fields along $\sigma$ that satisfies:

(i) for all $C^\infty$ vector fields along $\sigma$, $\frac{D}{dt}(Y + Z) = \frac{D}{dt}Y + \frac{D}{dt}Z$

(ii) for all $f \in C^\infty[a, b]$ and all $C^\infty$ vector fields $Y$ along $\sigma$, $\frac{D}{dt}(fY) = fY + f \cdot \frac{D}{dt}Y$

(iii) If $t_0 \in [a, b]$ and $\mathbf{Y}$ is a smooth vector field defined on a neighborhood of $\sigma(t_0) \in M$ then

$$\frac{D}{dt}|_{t_0} Y = \nabla_{\sigma'(t_0)} \mathbf{Y}.$$  

(305)

Definition 4.20 The unique operation $\frac{D}{dt}$ characterized in proposition (??) is called the covariant derivative along the curve $\sigma$.

Proof (of proposition (??)). The easiest way is to verify the proposition is to use a coordinate chart to obtain an extension $\mathbf{Y}$ of $Y$ (which depends on the coordinates).

Thus suppose $\sigma : [a, b] \mapsto M$ is a smooth curve, $Y : [a, b] \mapsto TM$ is a smooth vector field along $\sigma$ and $p = \sigma(t)$ for some $t \in [a, b]$. Pick any chart $(u, U)$ about $p$. Expand $Y$ in terms of the coordinate vector fields evaluated along the curve, i.e. $Y(t) = \sum_{j=1}^{m} (Y_{t} u^j) \frac{\partial}{\partial u^j}|_{\sigma(t)}$. Note that each map $t \mapsto \frac{\partial}{\partial u^j}|_{\sigma(t)}$ is a smooth vector field along (a segment of) the curve $\sigma$. Now use the properties (i), (ii), and (iii) to calculate:

$$\frac{DY}{dt}|_t = \frac{m}{\sum_{j=1}^{m} \left( (Y_{t} u^j)'(t) \frac{\partial}{\partial u^j}|_{\sigma(t)} + (Y_{t} u^j) \frac{D}{dt}|_t \frac{\partial}{\partial u^j}|_{\sigma(t)} \right)}$$

$$= \frac{m}{\sum_{j=1}^{m} \left( (Y_{t} u^j)'(t) \frac{\partial}{\partial u^j}|_{\sigma(t)} + (Y_{t} u^j) \nabla_{\sigma'(t)} \frac{\partial}{\partial u^j} \right)}$$

$$= \frac{m}{\sum_{j=1}^{m} \left( (Y_{t} u^j)'(t) \frac{\partial}{\partial u^j}|_{\sigma(t)} + (Y_{t} u^j) \sum_{i=1}^{m} (\sigma_i'(t)) \frac{\partial}{\partial u^j} \nabla_{\sigma'(t)} \frac{\partial}{\partial u^j} \right)}$$

$$= \sum_{k=1}^{m} \left( (Y_{t} u^k)'(t) + \sum_{i,j=1}^{m} \Gamma_{ij}^k(\sigma(t)) \cdot (\sigma_i'(t)) \cdot (Y_{t} u^j) \right) \frac{\partial}{\partial u^k}|_{\sigma(t)}$$

(306)

This calculation shows that there is no choice, i.e. given a connection $\nabla$, the conditions (i) through (iii) uniquely determine the value of $\frac{DY}{dt}$. We leave it as an exercise that the $\frac{D}{dt}$ is indeed well-defined, i.e. does not depend on the choice of the chart $(u, U)$. 

Exercise 4.72 Suppose $p$, $\sigma$ and $Y$ are as in the proof of the proposition and suppose that $(u, U)$ and $(v, V)$ are charts about $p$. Verify that $\frac{DY}{dt}|_t$ is well-defined, i.e. its value does not depend on the choice of the chart in the preceding calculation.

Moreover, verify that the definition made does in fact have the properties (i), (ii) and (iii).

Note that even when $\sigma'(t) = 0$ for some $t \in [a, b]$ then $\frac{DY}{dt}|_t$ need not be zero! Indeed the double sum in the preceding calculation vanishes at such points, but the first term need not. In the case that $\sigma(t) = q$ is constant over an interval $[t_1, t_2]$ then the restriction $Y : [t_1, t_2] \mapsto T_qM$ is just a curve in the single tangent space at the point $q$, and the covariant derivative reduces to the derivative of a curve in $T_qM \cong \mathbb{R}^m$. 

Summarizing, a connection provides a notion of rate of change of one vector fields in the direction on another. An important special case is the implied notion of the rate of change of a vector field along a curve. This provides a means of comparing tangent vectors at different points, and suggests a notion of a constant vector field, and even provides for a natural notion of geodesics that does not require a Riemannian structure!

**Definition 4.21** A vector field \( Y \) along a curve \( \sigma \) is called parallel along \( \sigma \) if \( \frac{DY}{dt} \equiv 0 \).

**Definition 4.22** Suppose \( \nabla \) is a connection with associated covariant derivative \( \frac{DY}{dt} \) along smooth curves. A smooth curve \( \sigma: [a,b] \to M \) is called a geodesic if it satisfies \( \frac{DY}{dt} \sigma' \equiv 0 \)

**Exercise 4.73** Verify that in the case of a Riemannian manifold with Levi-Civita connection \( \nabla \) this notion of geodesic agrees with that as a critical point of the energy functional.

**Proposition 4.44** Suppose \( \sigma: [a,b] \to M \) is a smooth curve and \( Y_a \in T_{\sigma(a)}M \). Then there exists a unique vector field \( Y: [a,b] \to TM \) along \( \sigma \) that is parallel along \( \sigma \).

**Proof.** Consider (25) with \( \frac{DY}{dt} \equiv 0 \) as a differential equations for \( Y \). In each chart \( (u,U) \) this is a linear equation, and hence the initial value problem has a unique, globally defined solution, i.e. on a maximal time interval \( (t_1,t_2) \) such that the image \( \sigma((t_1,t_2)) \subseteq U \) is a connected subset of \( U \). Since \( \sigma([a,b]) \subseteq M \) is compact it suffices to consider a finite sequence of charts, and it is obvious that \( Y \) extends to a unique vector field along \( \sigma \) defined on all of \( [a,b] \).

**Definition 4.23** The vector field \( Y \) defined in proposition (25) is called the parallel transport of \( Y_a \in T_{\sigma(a)}M \) along \( \sigma \). It is convenient to write \( \tau_t(Y_a) \in T_{\sigma(t+s)}M \) for the value of the parallel transport of \( Y_a \in T_{\sigma(s)}M \) along \( \sigma \), i.e., technically along the restriction of \( \sigma \) to \([s,s+t] \) or \([s+t,s]\). (This makes sense as long as \( s,t,s+t \in [a,b] \).)

Note that with this definition the domain of \( \tau_t \) is the union \( \bigcup_s T_{\sigma(s)}M \) taken over all those \( s \in [a,b] \) such that \( s+t \in [a,b] \). Since in most cases confusion is unlikely, one usually omits precise descriptions of the domain . . .

**Proposition 4.45** Suppose that \( I \subseteq \mathbb{R} \) is an interval, \( \sigma: I \to M \) is a smooth curve with associated parallel transport \( \tau \). Then the following hold for all \( t,s \in I \) such that \( t+s \in I \)

(i) On the respective domains \( \tau_s \circ \tau_t = \tau_{s+t} \)

(ii) For each \( t \) the restriction of \( \tau_t \) to \( T_{\sigma(s)}M \) is an isomorphism onto \( T_{\sigma(s+t)}M \).

(iii) If \( \nabla \) is the Levi-Civita connection on a Riemannian manifold \( M \), then the restriction of \( \tau_t \) to \( T_{\sigma(s)}M \) is an isometry.

**Proof.** The first property is obvious since \( \tau \) is defined as the flow of a linear differential equation (on a subset of \( TM \)). Taking \( s = -t \) shows that \( \tau_t \circ \tau_{-t} \) is the identity (on the appropriate union of tangent spaces). Hence each \( \tau_t \) is a bijection.

Now suppose \( M \) is a Riemannian manifold and \( \nabla \) the Levi-Civita connection on \( M \). In view of the polarization identity \( \langle a, b \rangle = \frac{1}{2}([a+b]^2 - \|a-b\|^2) \) it suffices to show that parallel transport is norm-preserving. Thus suppose \( \sigma: [a,b] \to M \) a smooth curve and \( Y \) a parallel vector field along \( \sigma \). Note that \( \frac{DY}{dt} \) is again a smooth vector field along a curve, and consequently \( \|Y\|^2: [a,b] \to \mathbb{R} \) is a continuous function. As a consequence of the subsequent lemma it is clear that \( \|Y\|^2 \) is constant on any interval \((t_1,t_2)\) on which \( \sigma' \) is nonzero. By continuity it follows that \( \|Y\|^2 \) is also constant on any interval on which \( \sigma' \) vanishes only at isolated points.
If $\sigma'(t) = 0$ for all $t$ in some interval $[t_1, t_2]$ then the restriction of $\sigma$ and $Y$ to this subinterval are a constant curve $\sigma(t) \equiv q$ and a curve in $T_pM$. But as is clear from equation (307), $Y$ must be constant: since $Y$ is parallel along $\sigma$ the left hand side vanishes, and since $\sigma' = 0$ the double sum is zero. Again invoking continuity, since $\|Y\|^2$ is constant on either kind of subinterval, it is constant throughout. 

\begin{lemma}
Suppose $Y$ and $Z$ are smooth vector fields along a smooth curve $\sigma: (-\varepsilon, \varepsilon) \mapsto M$ with $p = \sigma(0)$ and $X_p = \sigma'(0) \neq 0$. Then for all $t_0$ sufficiently close to $t = 0$
\begin{equation}
\left. \frac{d}{dt} \right|_{t_0} \langle Y, Z \rangle = \left. \left( \frac{D}{dt} \right|_{t_0} Y, Z \right\rangle + \langle Y_{t_0}, \left. \frac{D}{dt} \right|_{t_0} Z \rangle.
\end{equation}
\end{lemma}

**Proof.** Suppose $\sigma, Y, Z$ and $X_p$ are as in the statement of the lemma. Since $\sigma'(0) \neq 0$ the curve is locally an immersion, and thus there exist smooth vector fields $\overline{Y}, \overline{Z}, \in C^\infty(p)$ such that $Y = \overline{Y} \circ \sigma$ and $Z = \overline{Z} \circ \sigma$ (in some interval about 0). From the definitions it follows that
\begin{equation}
\left. \frac{d}{dt} \right|_{t_0} \langle Y, Z \rangle = \left. \frac{d}{dt} \right|_{t_0} \langle \overline{Y} \circ \sigma, \overline{Z} \circ \sigma \rangle = \langle \nabla_{X_p} \overline{Y}, \overline{Z}_p \rangle + \langle \overline{Y}_p, \nabla_{X_p} \overline{Z} \rangle = \langle \left. \frac{D}{dt} \right|_{t_0} Y, Z_{t_0} \rangle + \langle Y_{t_0}, \left. \frac{D}{dt} \right|_{t_0} Z \rangle.
\end{equation}
The foregoing has defined the parallel transport along a curve in terms of the covariant derivative along a curve, i.e. in terms of a connection $\nabla$. But one may proceed in the opposite way, starting with a notion of parallel transport to define a connection. As a step in that direction the following proposition shows that from the parallel transport as defined above one can recover the connection:

\begin{proposition}
Suppose $\sigma: (-\varepsilon, \varepsilon) \mapsto M$ is a smooth curve with $\sigma(0) = p$ and $\sigma'(0) = X_p$, and $Y \in \Gamma^\infty(p)$ is a smooth vector field defined in a neighborhood of $p$ containing $\sigma(-\varepsilon, \varepsilon)$. Then
\begin{equation}
\nabla_{X_p} Y = \lim_{h \to 0} \frac{1}{h} (\tau_{-h} Y_{\sigma(t)} - Y_p).
\end{equation}
\end{proposition}

**Proof.** Recall from exercise (306) that for a vector field $Y \in C^\infty(p)$ the value of $\langle \nabla_{X} Y \rangle_p$ is completely determined by the restriction of $Y$ to any curve $\sigma$ such that $\sigma'(0) = X_p$. Thus the proof reduces to a simple calculation when the given vector field $Y$ along the curve $\sigma$ is expanded as a linear combination of parallel vector fields. Choose a basis $\{W_{10}, W_{20}, \ldots, W_{m0}\} \subseteq T_pM$ and let $W_1, W_2, \ldots, W_m: (-\varepsilon, \varepsilon) \mapsto TM$ denote their parallel transports along $\sigma$. Since each $\tau_t$ is an isometry, it is clear that $\{W_{1t}, W_{2t}, \ldots, W_{mt}\} \subseteq T_{\sigma(t)}M$ form a basis for each $t$. Consequently, there exist functions $a^i: (-\varepsilon, \varepsilon) \mapsto \mathbb{R}$ such that $Y = \sum_{i=1}^{m} a^i W_i$. Since each $W_i$ is parallel along $\sigma$ it follows that $\lim_{h \to 0} \frac{1}{h} (\tau_{-h} Y_{\sigma(t)} - Y_p) = \sum_{i=1}^{m} (a^i(t) - a^i(0)) W_{i0}$ and hence
\begin{equation}
\lim_{h \to 0} \frac{1}{h} (\tau_{-h} Y_{\sigma(t)} - Y_p) = \sum_{i=1}^{m} \left. \frac{d}{dt} \right|_{t=0} a^i \cdot W_i(0) = \left. \frac{D}{dt} \right|_{t=0} \sum_{i=1}^{m} a^i \cdot W_i = \nabla_{X_p} Y.
\end{equation}

This construction is very similar to that of the Lie derivative which used the tangent map of the flow in place of the parallel transport employed here. The similarities suggest that one may in analogy define covariant derivatives of tensor fields in terms of the parallel transport. This is indeed possible – for details see e.g. Spivak II, pp.252–253.
Exercise 4.74 Explore defining the covariant derivative $\nabla_X \omega$ of a differential one-form $\omega$ in terms of the parallel transport along a curve. Use the result to compare $X(\omega(Y))$ with $\omega(\nabla_X Y)$ and $(\nabla_X \omega)(Y)$. For a function $f \in C^\infty(M)$, how does $\nabla_X (df)$ compare to $d(\nabla_X f) = d(Xf)$?

Exercise 4.75 Make a table comparing and contrasting Lie derivatives and the tangent maps of the flow on one side, and connections and parallel transport on the other side. Pay special attention to the (lack of) linearity over functions $f \in C^\infty(M)$, to which vector fields may be replaced by tangent vectors, to extensibility to tensor fields, to invariant definitions and descriptions in local coordinates, to the special case of (coordinate) fields in Euclidean spaces, ... As the next section shows, the integrability conditions on one side will be contrasted by conditions for flatness and characterizations of curvature.