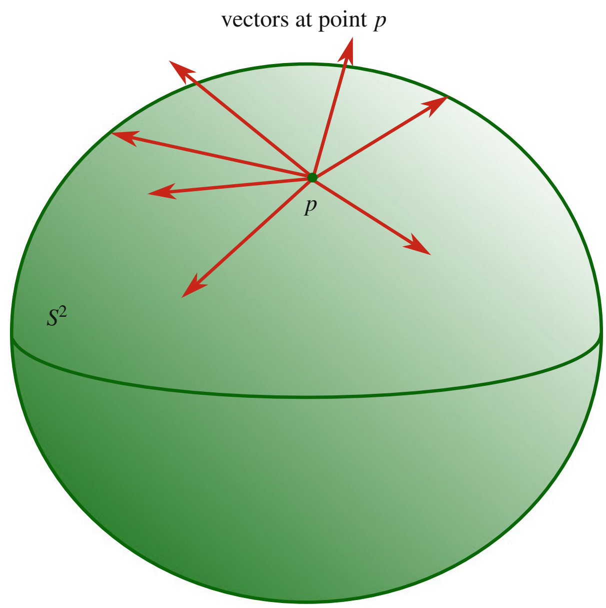

Fig. 10.9 The manifold S^2 along with a number of vectors at the point p tangent to S^2. However, these vectors are not actually in S^2. They are in fact in $\mathbb{R}^3$, the space in which we embed S^2 in order to draw it

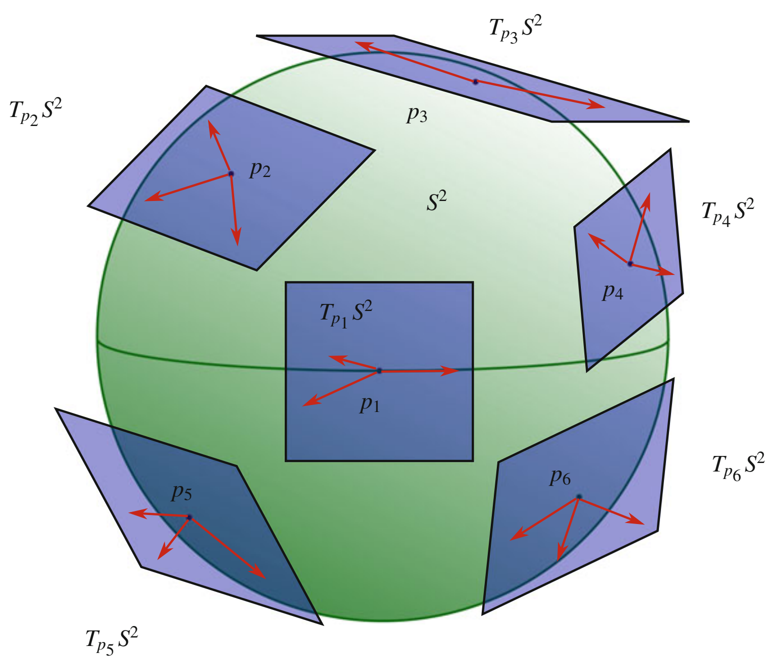

Fig. 10.10 When we introduced the concept of tangent spaces to the manifold S^2 we drew tangent planes to the manifold at various point on the manifold. Every point of S^2 actually has such a tangent plane

This is a perfectly good cartoon picture to help us visualize and understand tangent spaces, and we will continue using it, but it requires that our manifold be embedded in some $\mathbb{R}^n$ in which the tangent space is also embedded, just like both S^2 and the tangent spaces of S^2 are embedded in $\mathbb{R}^3$. According to a deep theorem in differential geometry called the Whitney embedding theorem any reasonably nice manifold can be embedded into $\mathbb{R}^n$ for some sufficiently large n, which means that we can actually continue to use our cartoon picture to help us visualize and think about tangent spaces. However, we essentially have to "step outside" of the manifold in order to visualize tangent spaces this way. A property that requires us to "step outside" of the manifold in order to think about it or define it is called an **extrinsic** property.

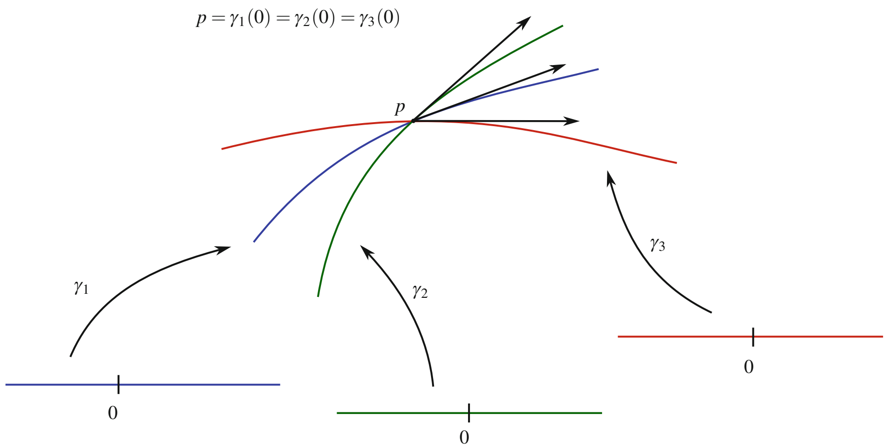

Fig. 10.11 Three curves $\gamma_i : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$, $i = 1, 2, 3$, such that $\gamma_i(0) = p$. Each curve has a tangent vector at the point p

We would like to be able to define and think about tangent spaces without "stepping outside" the manifold. There are many cases where being able to think about manifold properties while still "being inside" the manifold is useful. A property of a manifold that we can define or think about while still "being inside" the manifold is call an **intrinsic** property. An intrinsic property does not, in any way, require or rely on the manifold being embedded into some $\mathbb{R}^n$. Even though we have been thinking of tangent vectors and tangent spaces extrinsically up to now they are, in fact, intrinsic properties. We can define them and think about them while still "being inside" the manifold. When we do this the idea of arrows emanating from a point is no longer appropriate.

We will approach the intrinsic definition of tangent vectors by first considering all **smooth curves**

$$\gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$$

such that $\gamma(0) = p \in M$ and ϵ is just some small positive number. In Fig. 10.11 we draw several smooth curves $\gamma_i : (-\epsilon, \epsilon) \to M$, where $\gamma_i(0) = p$ for all i. We will call the parameter that γ_i depends on time. Without getting into the technical meaning of a smooth curve just assume that it means what you intuitively think it means, that there are no sharp corners on the curve.

Each tangent vector v_p is actually tangent to some smooth curve that goes through the point $p \in M$. But while the tangent vector v_p may actually only exist in the space $\mathbb{R}^n$ in which M is embedded, the curves γ exist entirely in the manifold M. In essence each of these curves defines a different tangent vector at the point p. Thus the general idea is to identify tangent vectors to the manifold M at a point p with curves in the manifold that go through the point p, thereby eliminating our need to embed M into some larger space $\mathbb{R}^n$ in order to have tangent vectors.

It may seem possible that the tangent space of M is equivalent to the set of curves $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$. This is almost true, but not quite. There are two issues we have to consider carefully. To address the first issue consider Fig. 10.12. Here we have three different curves $\gamma_1, \gamma_2, \gamma_3 : (-\epsilon, \epsilon) \to M$ with $\gamma_1(0) = \gamma_2(0) = \gamma_3(0) = p$ that are identical very close to p. If two curves are identical very close to p then can we in some sense consider them the same curve? At a first glance it would seem reasonable to do this.

To consider the second issue, consider the vectors v_p and $2v_p$ in T_pM. From Fig. 10.13 it looks like both v_p and $2v_p$ can be represented by the same curve γ. The question is, when we use curves how would we distinguish between the vectors v_p and $2v_p$ in T_pM? They are both pointing in the same direction, but the second vector is twice as long as the first vector.

The clue to how we would handle this second issue is that we called the parameter that γ depends on time. The derivative of γ at the time $t = 0$ gives a velocity at time 0. As we increase the rate at which we move along the curve γ then the velocity is greater and so the velocity vector becomes longer. So in order to distinguish between v_p and $2v_p$ the speed of the parametrization is increased so it is twice what it originally was at $t = 0$.

Let us consider a simple example. Let $M = \mathbb{R}^2_{xy}$ be the Euclidian plane parameterized by x and y. Consider the following three vectors at the origin of $\mathbb{R}^2_{xy}$:

$$u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad v_0 = 2u_0 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad w_0 = -u_0 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

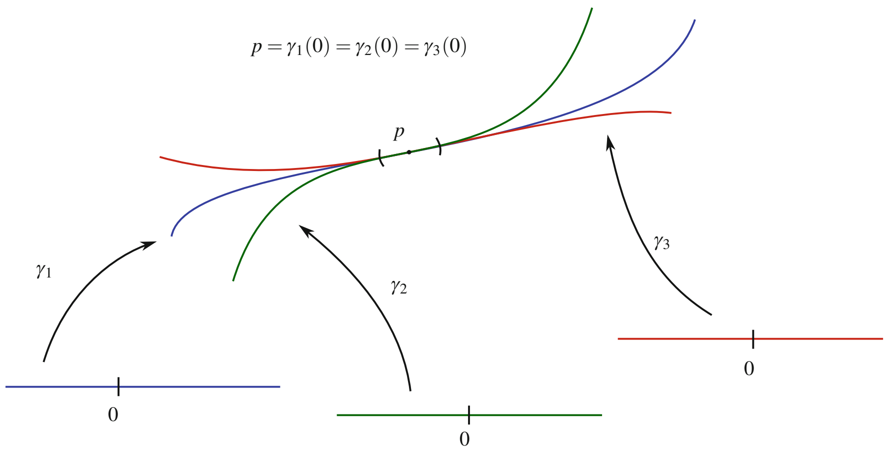

Fig. 10.12 As in Fig. 10.11 three curves $\gamma_i : (-\epsilon, \epsilon) \subset \mathbb{R} \longrightarrow M$, $i = 1, 2, 3$, such that $\gamma_i(0) = p$. However, here when we are very close to p these curves are identical

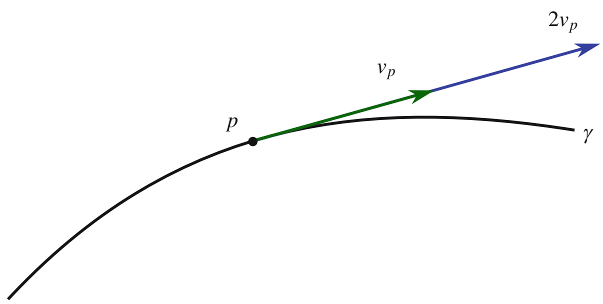

Fig. 10.13 Two vectors, v_p and $2v_p$, that both appear to be tangent to the same curve γ

See Fig. 10.14. These vectors are given by the following three curves $\alpha, \beta, \gamma : (-\epsilon, \epsilon) \to \mathbb{R}^2_{xy}$ where

$$\begin{aligned} \alpha(t) &= (\alpha_1(t), \alpha_2(t)) = (t, 2t), \\ \beta(t) &= (\beta_1(t), \beta_2(t)) = (2t, 4t), \\ \gamma(t) &= (\gamma_1(t), \gamma_2(t)) = (-t, -2t). \end{aligned}$$

Though drawing the curves α, β, and γ in $\mathbb{R}^2$ will give what looks like the same curve, technically they are all different curves because of their different parameterizations.

Now we will define the tangent space of M at point p, T_pM, in terms of curves on M, but we have to take the two above issues into consideration. Two curves that are identical very close to p, as in Fig. 10.12, have the same range close to p. If two curves have the same range close to p and have the same parametrization close to p they are called equivalent, which is denoted $\sim$. Suppose that γ_1, γ_2, and γ_3 all have the same range close to p and all have the same parametrization close to p then we would say $\gamma_1 \sim \gamma_2 \sim \gamma_3$. The set of all equivalent curves is called an **equivalence class**. The equivalence class of γ_1 is the set of all the curves equivalent to γ_1, is denoted by $[\gamma_1]$, and is defined by

$$[\gamma_1] \equiv \big\{\gamma \;\big|\; \gamma \sim \gamma_1\big\}.$$

Thus, if $\gamma_1 \sim \gamma_2 \sim \gamma_3$ we would have $[\gamma_1] = [\gamma_2] = [\gamma_3]$. This means that which member of the equivalence class you use to represent the equivalence class does not matter. Each equivalence class of curves at a point p is defined to be a tangent vector at p. We will write $[\gamma]_p$ when we want to indicate that the base point is p. The **tangent space** of M at p defined as the set of all tangent vectors, that is, equivalence classes of curves, at the point p,

$$T_pM = \Big\{ [\gamma]_p \;\Big|\; \gamma : (-\epsilon, \epsilon) \to M \text{ and } \gamma(0) = p \Big\}.$$

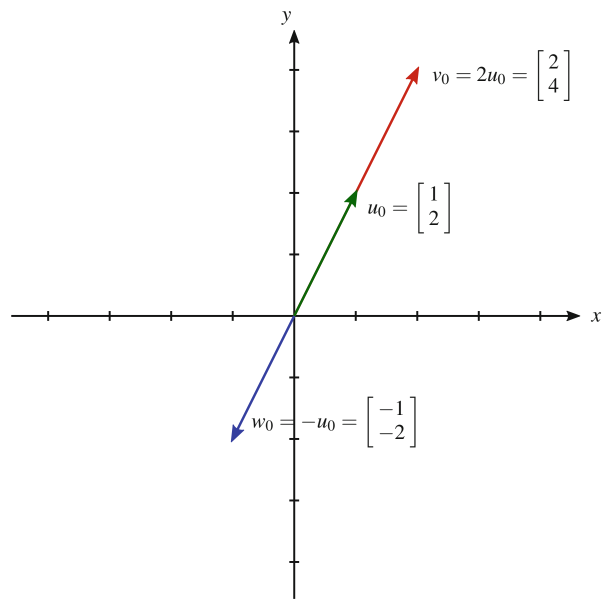

Fig. 10.14 The manifold $\mathbb{R}^2_{xy}$ with three vectors $u_0 = [1, 2]^T$, $v_0 = 2u_0 = [2, 4]^T$, and $w_0 = -u_0 = [-1, -2]^T$ at the origin. These three vectors are determined by three curves α, β, and γ with different parameterizations

Thus the tangent space to M at the point p is the set of all equivalence classes of curves that go through $p \in M$. Notice that this definition of the tangent space is intrinsic, it entirely relies on curves in the manifold M that go through the point p. Thus we do not need to "step outside" our manifold in order to visualize or think about the tangent space. This definition is independent of any bigger space that our manifold is embedded in.

Now we want to understand the relationship between our extrinsic and intrinsic views of vectors and tangent spaces. A vector v_p, which actually lies in the space our manifold is embedded in, is the velocity vector of some curve γ at the point p so we have made the identification $v_p = [\gamma]_p$ To obtain back the vector v_p from the curve γ we have to take the derivative of γ with respect to time and evaluate the derivative at time $t = 0$. Basically, we are finding the Jacobian matrix of the mapping $\gamma : (-\epsilon, \epsilon) \to M$.

Returning to our concrete example we have

$$u_{(0,0)} \equiv [\alpha]_{(0,0)} \quad v_{(0,0)} \equiv [\beta]_{(0,0)}, \quad w_{(0,0)} \equiv [\gamma]_{(0,0)}.$$

But notice, even though when we draw the curves on $\mathbb{R}^2$ and it turns out that they all look like "the same" curve in our picture, they are really different curves because they have different parameterizations and so are not part of the same equivalence class of curves. In other words, we have $\alpha \not\sim \beta$, $\alpha \not\sim \gamma$, and $\beta \not\sim \gamma$. Usually we simply use 0 to represent the base point $(0, 0)$ but since we want to emphasize the difference between the base point and evaluating the partial derivatives at time $t = 0$ we will be extra careful with our notation here. We can see how our two views of vectors correlate,

$$u_{(0,0)} \equiv \Big[\big(\alpha_1(t), \alpha_2(t)\big)\Big]_{(0,0)} = \big[(t, 2t)\big]_{(0,0)},$$

$$u_{(0,0)} = \begin{bmatrix} \left.\frac{\partial \alpha_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial \alpha_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \left.\frac{\partial (t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial (2t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{(0,0)}.$$

Similarly we have,

$$v_{(0,0)} \equiv \Big[\big(\beta_1(t), \beta_2(t)\big)\Big]_{(0,0)} = \big[(2t, 4t)\big]_{(0,0)},$$

$$v_{(0,0)} = \begin{bmatrix} \left.\frac{\partial \beta_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial \beta_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \left.\frac{\partial (2t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial (4t)}{\partial t}\right|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}_{(0,0)},$$

and

$$w_{(0,0)} \equiv \Big[\big(\gamma_1(t), \gamma_2(t)\big)\Big]_{(0,0)} = \big[(-t, -2t)\big]_{(0,0)},$$

$$w_{(0,0)} = \begin{bmatrix} \frac{\partial \gamma_1(t)}{\partial t}\Big|_{t=0} \\ \frac{\partial \gamma_2(t)}{\partial t}\Big|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} \frac{\partial(-t)}{\partial t}\Big|_{t=0} \\ \frac{\partial(-2t)}{\partial t}\Big|_{t=0} \end{bmatrix}_{(0,0)} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}_{(0,0)}.$$

Below is a key for moving between the two ways we have defined vectors on a manifold. The equivalence class of curves definition is based on a curve that exists entirely inside our manifold M and does not require M to be embedded into a larger ambient space in order to "see" the vectors; our vectors are simply viewed as the curves that lie entirely in M. What I will call here the vector definition requires either embedding our manifold M into an ambient space $\mathbb{R}^n$ or drawing the tangent space to the manifold M at a point in order to "see" the vector. **In the equivalence class definition we view the curve as the vector and in the vector definition we view the derivative of the curve as the vector.** This may seem odd to you but it is simply thinking about vectors from two different perspectives. Hopefully this will not be too confusing.

Equivalence Class Definition	$\Longleftrightarrow$	Vector Definition
$v_p = [\gamma]_p, \quad p = \gamma(0)$ $= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]$		$v_p = [\gamma']_p = [\gamma]'_p$ $= \Big[\big(\gamma'_1(t), \ldots, \gamma'_n(t)\big)\Big]_p$ $= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]'_p$ $= \begin{bmatrix} \gamma'_1(t=0) \\ \vdots \\ \gamma'_n(t=0) \end{bmatrix}_p$ $= \begin{bmatrix} \frac{\partial \gamma_1(t)}{\partial t}\Big\|_{t=0} \\ \vdots \\ \frac{\partial \gamma_n(t)}{\partial t}\Big\|_{t=0} \end{bmatrix}_p$

One more comment is in order here. As we know, integration is in a sense the opposite of differentiation. If we can view a vector v_p as the derivative of a curve γ at the point p, can we view the curve γ as the integral of the vector v_p? The problem here is that v_p exists at only one single point, and the idea of integration at a single point is ill-defined. What we need is a **smooth vector field** around a point in order to do integration. Without getting into the technical details of what a smooth vector field is, think of it as a vector field which changes smoothly as you vary the point. Once we have a smooth vector field in some neighborhood then we can talk of curves called **integral curves**. The curve γ that is used to represent the equivalence class $[\gamma]_p$ of some vector v_p need not be an integral curve, but if v_p is an element of a smooth vector field then usually people make the unstated assumption that γ is an integral curve. We will discuss integral curves in more detail in the contexts of electromagnetism and geometric mechanics in Chap. 12 and Appendix B.

Recall that in Sect. 2.3 we equated tangent vectors at a point p with directional derivatives. By doing this we identified the Euclidian basis vectors of the tangent space with partial derivative operators. In other words, we had found the following identifications,

$$e_{1_p} = \frac{\partial}{\partial x}\bigg|_p, \quad e_{2_p} = \frac{\partial}{\partial y}\bigg|_p, \quad e_{3_p} = \frac{\partial}{\partial z}\bigg|_p.$$

Thus, for a vector $v_p \in T_p\mathbb{R}^3$,

$$v_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_p$$

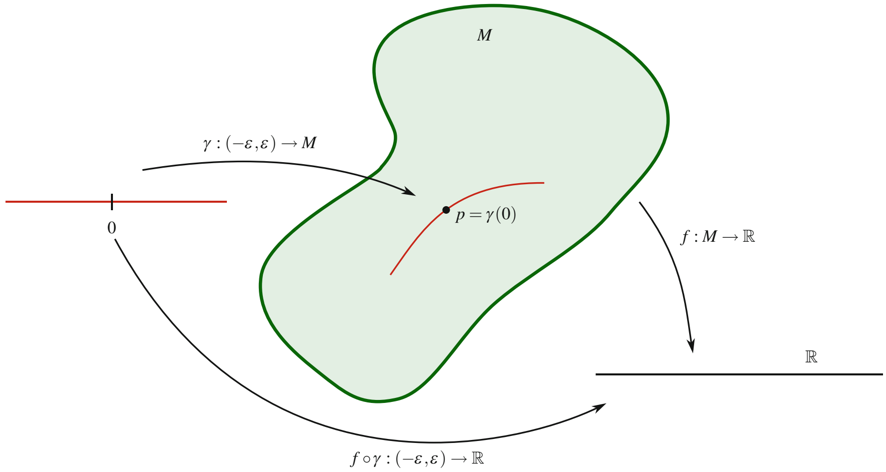

Fig. 10.15 The vector $[\gamma]_p$ given by $\gamma : (-\epsilon, \epsilon) \to M$, where $\gamma(0) = p$, and a function $f : M \to \mathbb{R}$. Then we have $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$, which is the sort of mapping we know how to differentiate

$$
\begin{aligned}
&= v_1 e_{1_p} + v_2 e_{2_p} + v_3 e_{3_p} \\
&= v_1 \left.\frac{\partial}{\partial x}\right|_p + v_2 \left.\frac{\partial}{\partial y}\right|_p + v_3 \left.\frac{\partial}{\partial z}\right|_p .
\end{aligned}
$$

The directional derivative of the function $f : \mathbb{R}^3 \to \mathbb{R}$ in the direction v_p was written as the vector v_p operating on the function f,

$$
\begin{aligned}
v_p[f] &= \left(v_1 \left.\frac{\partial}{\partial x}\right|_p + v_2 \left.\frac{\partial}{\partial y}\right|_p + v_3 \left.\frac{\partial}{\partial z}\right|_p \right)[f] \\
&= v_1 \left.\frac{\partial f}{\partial x}\right|_p + v_2 \left.\frac{\partial f}{\partial y}\right|_p + v_3 \left.\frac{\partial f}{\partial z}\right|_p .
\end{aligned}
$$

We want to consider what this is in our more abstract intrinsic setting where the vector v_p is given by an equivalence class of curves $[\gamma]_p$, that is, when we have $v_p \equiv [\gamma]_p$. Before we used the notation $v_p[f]$ to denote $D_{v_p} f$, directional derivative of f in the direction of v_p. We used this notation to emphasise the fact that it is the vector v_p which is, in a sense, operating on the function f. We will continue to use this notation whenever we write the vector as v_p, but when we write the vector using our equivalence class of curves notation $[\gamma]_p$ then writing $[\gamma]_p[f]$ would be a little odd since the two sets of square brackets $[\cdot]$ have different meanings. Notation is sometimes as much the result of habit and convention as rational planning.

Consider $\gamma : (-\epsilon, \epsilon) \to M$, where $\gamma(0) = p$, and $f : M \to \mathbb{R}$. Then $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$, as shown in Fig. 10.15. When our vector v_p is given by an equivalence class of curves $[\gamma]_p$ then the directional derivative of f in the v_p direction is given by

$$
\begin{aligned}
v_p[f] &= D_{[\gamma]_p} f \\
&= [f \circ \gamma]'_p \\
&= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_{t=0} \\
&= \lim_{h \to 0} \frac{(f \circ \gamma)(h) - (f \circ \gamma)(0)}{h}.
\end{aligned}
$$

One often used notation is $D_{[\gamma]_p} f$, which mirrors the notation generally used in multi-variable calculus. Other notations for $D_{[\gamma]_p} f$ include $[(f \circ \gamma)']_p$ or $[f \circ \gamma]'_p$ or $[(f \circ \gamma)']_{t=0}$ or even $[f \circ \gamma]'_{t=0}$. There is no perfectly standard way of writing this and one can use all sorts of different permutations in the notation. The key of course is that if the vector is at point

$p = \gamma(0)$ that means that the derivative of γ has to be evaluated at $t = 0$. This is the pertinent point. Also, usually one just writes a 0. You need to pay attention to context to decide whether the 0 means to evaluate at $t = 0$ or the base point is the origin $(0, 0)$. With this warning made we will stop trying to be so precise.

Notice that the directional derivative of f at the point $p = \gamma(0)$ in the direction $[\gamma]_p$ is defined to be the same thing as the ordinary one-variable derivative of the function $f \circ \gamma$ at $t = 0$. It is simple enough to see that we arrive at the same answer with this notation. Assume $\gamma(t) = \big(x(t), y(t), z(t)\big)$, then we have

$$\begin{aligned} v_p &= \Big[\big(x(t), y(t), z(t)\big)\Big]'_p \\ &= \begin{bmatrix} x'(t=0) \\ y'(t=0) \\ z'(t=0) \end{bmatrix}_p . \end{aligned}$$

Using $f \circ \gamma(t) = f\big(x(t), y(t), z(t)\big)$ and we then get

$$\begin{aligned} v_p[f] &= D_{[\gamma]_p} f \\ &= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_{t=0} \\ &= \left.\frac{\partial\ f\big(x(t), y(t), z(t)\big)}{\partial t}\right|_{0} \\ &= \left.\frac{\partial f}{\partial x}\right|_{\gamma(0)} \cdot \left.\frac{\partial x}{\partial t}\right|_{0} + \left.\frac{\partial f}{\partial y}\right|_{\gamma(0)} \cdot \left.\frac{\partial y}{\partial t}\right|_{0} + \left.\frac{\partial f}{\partial z}\right|_{\gamma(0)} \cdot \left.\frac{\partial z}{\partial t}\right|_{0} \qquad \text{chain rule} \\ &= \left.\frac{\partial f}{\partial x}\right|_{p} \underbrace{x'(0)}_{v_1} + \left.\frac{\partial f}{\partial y}\right|_{p} \underbrace{y'(0)}_{v_2} + \left.\frac{\partial f}{\partial z}\right|_{p} \underbrace{z'(0)}_{v_3}, \end{aligned}$$

which is exactly the same as what we had for $v_p[f]$ just above, with the base point p left off the notation.

Turning back to a concrete example, suppose we had the function $f : \mathbb{R}^2_{xy} \to \mathbb{R}$ given by $f(x, y) = x + \sin(y)$. We want to find $u_0[f]$, where $u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \big[(t, 2t)\big]_0$. (Note, here we have $0 = (0, 0)$.) Proceeding as we would have before we have

$$\begin{aligned} u_0[f] &= \left(1 \cdot \left.\frac{\partial}{\partial x}\right|_0 + 2 \cdot \left.\frac{\partial}{\partial y}\right|_0\right) f \\ &= 1 \cdot \left.\frac{\partial\big(x + \sin(y)\big)}{\partial x}\right|_0 + 2 \cdot \left.\frac{\partial\big(x + \sin(y)\big)}{\partial y}\right|_0 \\ &= 1 \cdot 1 + 2 \cdot \cos(0) \\ &= 3. \end{aligned}$$

However, with $u_0 = [\gamma]_0 = [(t, 2t)]_0$, we have

$$\begin{aligned} D_{u_0} f &= [f \circ \gamma]'_0 \\ &= \left.\frac{\partial (f \circ \gamma)}{\partial t}\right|_0 \\ &= \left.\frac{\partial\big(t + \sin(2t)\big)}{\partial t}\right|_0 \\ &= 1 + 2\cos(2t)\Big|_0 \\ &= 1 + 2 = 3. \end{aligned}$$

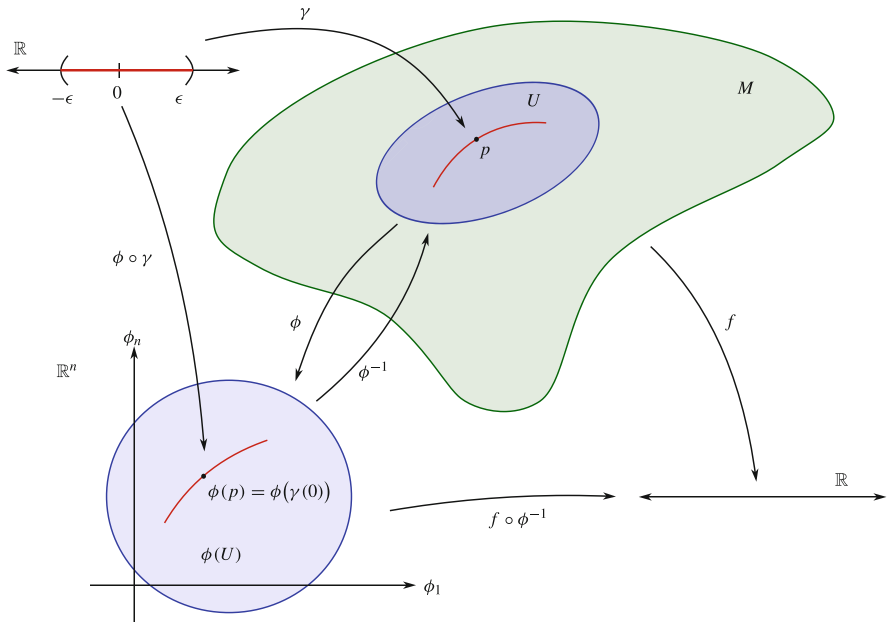

Fig. 10.16 Here the manifold M is shown with three things, (**a**) the curve γ that represents a tangent vector at the point p, (**b**) a real-valued function $f : M \to \mathbb{R}$ on the manifold M, and (**c**) a coordinate chart (U, ϕ) such that $p \in U$. It is obvious that $f \circ \gamma = f \circ \phi^{-1} \circ \phi \circ \gamma = (f \circ \phi^{-1}) \circ (\phi \circ \gamma)$. The mappings $\phi \circ \gamma : \mathbb{R} \to \mathbb{R}^n$ and $f \circ \phi^{-1} : \mathbb{R}^n \to \mathbb{R}$ are also shown

This all seems well and good, until you notice that in our example our manifold was $\mathbb{R}^2_{xy}$, that is, $\mathbb{R}^2$ with the Cartesian coordinate system "built in." Finding our composition $f \circ \gamma$ relied on knowing this coordinate system. How so? Since the function f was written explicitly in terms of the variables x and y which made writing down $f \circ \gamma$ easy. But how would we manage the directional derivatives on a manifold that does not have a nice standard coordinate system "built in"? In other words, what if it was not so straightforward to write down what $f \circ \gamma$ actually is?

If we want to find the derivative of $f : M \to \mathbb{R}$ at the point $p \in M$ in the direction of $v_p \equiv [\gamma]_p$ then we need to know how to compose f and γ to get the map $f \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \mathbb{R}$. But for an arbitrary manifold M this may not actually be clear. The composition $f \circ \gamma$ actually relies on some coordinate chart ϕ. Thus, our composition $f \circ \gamma$ must be done via some coordinate chart so that we can deal with maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, which we understand. Consider Fig. 10.16. Since differentiable manifolds come with coordinate charts (U, ϕ) we can make use of these. We have that

$$f \circ \gamma = f \circ \phi^{-1} \circ \phi \circ \gamma.$$

The map $\phi \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to \phi(U) \subset \mathbb{R}^n$ is the sort of map that we can actually do computations with and take derivatives of, and so is the map $f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n \to \mathbb{R}$. So we actually have

$$\begin{aligned} v_p[f] &= D_{[\gamma]_p} f \\ &= [f \circ \gamma]'_0 \\ &= [f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 \\ &= [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0 \\ &= \left.\frac{\partial (f \circ \phi^{-1})}{\partial \phi}\right|_{\phi(p)} \cdot \left.\frac{\partial (\phi \circ \gamma)}{\partial t}\right|_0 . \end{aligned}$$

Notice how we have

Chain Rule for curves	$[f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi \circ \gamma(0)} \cdot [\phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0.$

This is really one way of writing the chain rule. It may not look anything like the chain rule you are used to, but this is what the chain rule looks like in this context, when we are differentiating while using the equivalence-class-of-curves definition of vectors.

To show this let us consider each of these terms separately. First we look at the mapping

$$\begin{aligned}\phi \circ \gamma : (-\epsilon, \epsilon) \subset \mathbb{R} &\longrightarrow \phi(U) \subset \mathbb{R}^n \\ t &\longmapsto \Big(\phi_1\big(\gamma(t)\big), \phi_2\big(\gamma(t)\big), \ldots, \phi_n\big(\gamma(t)\big)\Big) \\ &= \Big(\phi_1 \circ \gamma(t), \phi_2 \circ \gamma(t), \ldots, \phi_n \circ \gamma(t)\Big)\end{aligned}$$

so $\left.\frac{\partial(\phi \circ \gamma)}{\partial t}\right|_0$ is really the Jacobian matrix evaluated at $t = 0$,

$$\left.\frac{\partial(\phi \circ \gamma)}{\partial t}\right|_0 = \begin{bmatrix} \frac{\partial(\phi_1 \circ \gamma)}{\partial t} \\ \frac{\partial(\phi_2 \circ \gamma)}{\partial t} \\ \vdots \\ \frac{\partial(\phi_n \circ \gamma)}{\partial t} \end{bmatrix}_0 .$$

Next we look at the mapping

$$\begin{aligned}f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\phi_1, \phi_2, \ldots, \phi_n) &\longmapsto f\big(\phi^{-1}(\phi_1, \phi_2, \ldots, \phi_n)\big) \\ & \qquad f \circ \phi^{-1}(\phi_1, \phi_2, \ldots, \phi_n)\end{aligned}$$

so $\left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi}\right|_{\phi(p)}$ is really the Jacobian matrix evaluated at $\phi(p)$,

$$\left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi}\right|_{\phi(p)} = \left[\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_2}, \ldots, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right]_{\phi(p)} .$$

Putting these two pieces together we have

$$\begin{aligned}D_{[\gamma]_p} f &= \left[\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_2}, \ldots, \frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right]_{\phi(p)} \begin{bmatrix} \frac{\partial(\phi_1 \circ \gamma)}{\partial t} \\ \frac{\partial(\phi_2 \circ \gamma)}{\partial t} \\ \vdots \\ \frac{\partial(\phi_n \circ \gamma)}{\partial t} \end{bmatrix}_0 \\ &= \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_1}\right|_{\phi(p)} \left.\frac{\partial(\phi_1 \circ \gamma)}{\partial t}\right|_0 + \cdots + \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_n}\right|_{\phi(p)} \left.\frac{\partial(\phi_n \circ \gamma)}{\partial t}\right|_0 \\ &= \sum_{i=1}^n \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_i}\right|_{\phi(p)} \left.\frac{\partial(\phi_i \circ \gamma)}{\partial t}\right|_0 .\end{aligned}$$

Also notice that this means we have

$$[f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0 = \sum_{i=1}^n \left.\frac{\partial(f \circ \phi^{-1})}{\partial \phi_i}\right|_{\phi(p)} \left.\frac{\partial(\phi_i \circ \gamma)}{\partial t}\right|_0 ,$$

which makes it clear that $[f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0$ is indeed simply another way of writing the chain rule.

Question 10.2 Suppose that (U, ϕ) and $(\tilde{U}, \tilde{\phi})$ are two charts on M, $p \in U \cap \tilde{U}$, $\gamma : (-\epsilon, \epsilon) \subset \mathbb{R} \to M$ with $\gamma(0) = p$, and f is a real-valued function on M. Show that

$$\left.\frac{\partial(f \circ \phi^{-1} \circ \phi \circ \gamma)}{\partial t}\right|_0 = \left.\frac{\partial(f \circ \tilde{\phi}^{-1} \circ \tilde{\phi} \circ \gamma)}{\partial t}\right|_0 .$$

10.3 Push-Forwards and Pull-Backs on Manifolds

We have already discussed push-forwards and pull-backs in Chap. 6. One of the examples we looked at closely was the change of variable formula $f : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $f(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$. We had found the push-forward of f to be the mapping

$$T_p f \equiv f_* \equiv D_p f : T_p\mathbb{R}^2_{xy} \longrightarrow T_{f(p)}\mathbb{R}^2_{uv}$$

given by the Jacobian matrix

$$T_p f \equiv f_* \equiv D_p f = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_p .$$

More generally, given a mapping between two manifolds $f : M \to N$ this mapping f induces another mapping $Tf : TM \to TN$ from the tangent bundle of M to the tangent bundle of N. At a particular point $p \in M$ we have $T_p f : T_p M \to T_p N$. This mapping pushes-forward vectors $v_p \in T_p M$ to vectors $T_p f \cdot v_p \in T_{f(p)}N$. See Fig. 10.17.

Now we want to look at this from our more abstract perspective. Suppose we had a map $\phi : M \to N$ and we had the vector v_p, which was determined by the curve γ. We want to know what $T_p\phi \cdot v_p$ is. From Fig. 10.18 it should be clear that the push-forward of vector $[\gamma]$ by the mapping ϕ is the curve $\phi \circ \gamma$. Thus we define the push-forward of $[\gamma]_p$ by $T_p\phi$ to be

Push-foraward of vector $[\gamma]_p$ by $T_p\phi$

$$T_p\phi \cdot [\gamma]_p = [\phi \circ \gamma]_{\phi(p)}.$$

In a lot of ways this is a much cleaner formulation of the push-forward. It also makes abstract computations much simpler, as we will see. Notationally there are several different ways to write this, all of which represent the same thing.

Ways to write push-foraward of vector $[\gamma]_p$by $T_p\phi$

$$\begin{aligned} T_p\phi \cdot [\gamma]_p &\equiv \phi_*(p)\big([\gamma]_p\big) \\ &\equiv \phi_*(p) \cdot [\gamma]_p \\ &\equiv \phi_*\big([\gamma]_p\big) \\ &\equiv [\phi \circ \gamma]_{\phi(p)}. \end{aligned}$$

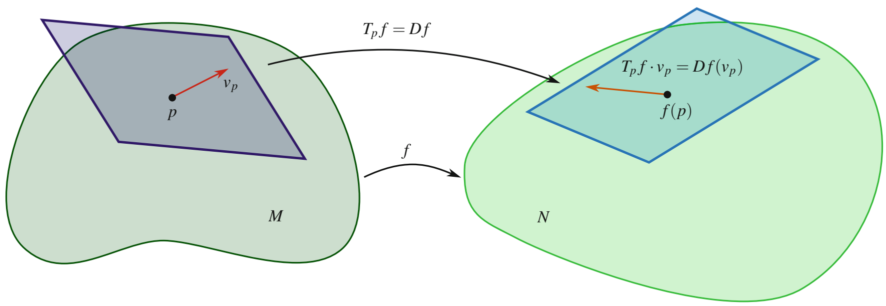

Fig. 10.17 Given a mapping $f : M \to N$ between manifolds then f induces the push-forward mapping $Tf : TM \to TN$. At a particular point we have $T_p f : T_p M \to T_p N$

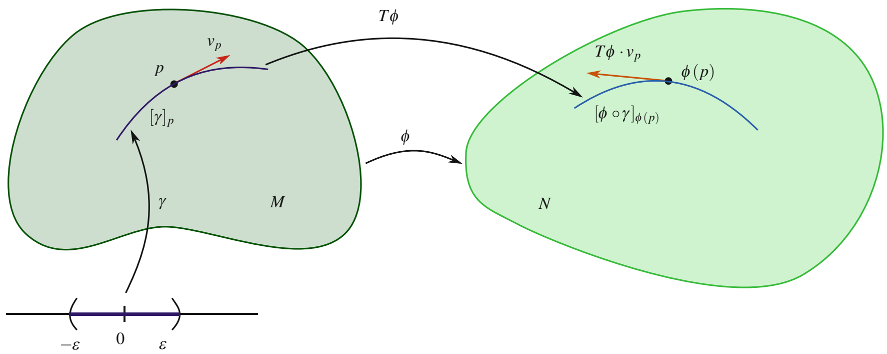

Fig. 10.18 Here we have a mapping $\phi : M \to N$ between manifolds and a vector $[\gamma]_p \in T_pM$, where $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$. What is the push-forward by $T_p\phi$ of this vector? It is simply the vector given by the curve $\phi \circ \gamma$, that is, by $[\phi \circ \gamma]_{\phi(p)}$

Of course, all of these notations can also be written leaving off the base point:

$$T\phi \cdot [\gamma] \equiv \phi_*\big([\gamma]\big) \equiv \phi_* \cdot [\gamma] = [\phi \circ \gamma].$$

Turning to our familiar example $\phi : \mathbb{R}^2_{xy} \to \mathbb{R}^2_{uv}$ given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = (x + y, x - y)$ we can see that these two ways of computing the push-forward of a vector give the same answer. Consider again the vector $u_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \equiv [\gamma]_0$ where

$$\begin{aligned} \gamma : (-\epsilon, \epsilon) &\longrightarrow \mathbb{R}^2_{xy} \\ t &\longmapsto (t, 2t). \end{aligned}$$

We wish to find the push-forward of u_0 by ϕ. First we do this by finding the tangent mapping $T\phi$ and then use that to find $T\phi \cdot u_0$.

$$T\phi = \phi_* = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We do not bother with worrying about the base point for this matrix since clearly $T\phi$ does not depend on the base point it is located at. We then find the push-forward of u_0 by

$$T\phi \cdot u_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\phi(0)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_0.$$

Now we compute the same thing using the more abstract equivalence class of curves definition of vectors,

$$T\phi \cdot [\gamma]_0 = [\phi \circ \gamma]_{\phi(0)} = [\phi \circ (t, 2t)]_0 = [(t + 2t, t - 2t)]_0 = [(3t, -t)]_0.$$

Thus the curve $\phi \circ \gamma = (3t, -t)$ is the push-forward vector we are looking for. That is, $T\phi \cdot u_0 \equiv [(3t, -t)]_0$. In order to check that our answer is correct and equivalent to what we found just above we have to take the derivative of this curve with respect to time,

$$[\phi \circ \gamma]'_{\phi(0)} = \begin{bmatrix} \left.\frac{\partial(\phi\circ\gamma)_1(t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial(\phi\circ\gamma)_2(t)}{\partial t}\right|_{t=0} \end{bmatrix}_{\phi(0)} = \begin{bmatrix} \left.\frac{\partial(3t)}{\partial t}\right|_{t=0} \\ \left.\frac{\partial(-t)}{\partial t}\right|_{t=0} \end{bmatrix}_{\phi(0)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_0,$$

which is exactly what we wanted to show.

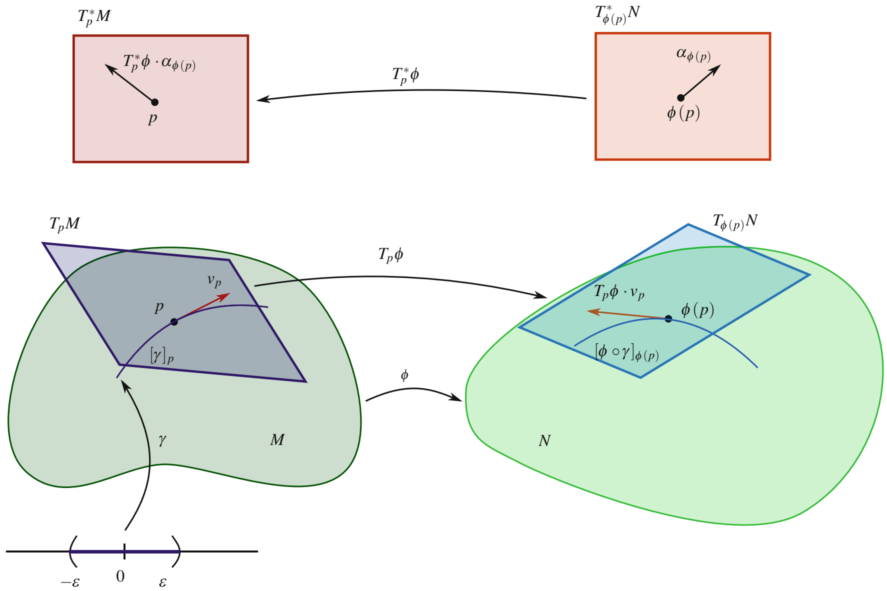

Fig. 10.19 Here the relationship between push-forward of vectors and pull-backs of one forms is illustrated. We picture the cotangent space T_p^*M and $T^*_{\phi(p)}N$ above the tangent space, even though both tangent space and cotangent space are "attached" to the manifold at the same point. Notice that the pull-back mapping goes in the opposite direction as the push-forward mapping

Question 10.3 For a mapping $\phi : M \to N$ and a function $f : N \to \mathbb{R}$, show that $v_p[f \circ \phi] = \big(T_p\phi \cdot v_p\big)_{\phi(p)}[f]$.

Now that we have addressed push-forwards of vectors we want to look at pull-backs of differential forms. Conceptually, what is happening is not very difficult to understand. The relevant mappings are defined as

$$\begin{aligned} \phi : M &\longrightarrow N \\ p &\longmapsto \phi(p) \end{aligned}$$

$$\begin{aligned} T_p\phi : T_pM &\longrightarrow T_{\phi(p)}N \\ v_p &\longmapsto T_p\phi \cdot v_p \\ [\gamma]_p &\longmapsto [\phi \circ \gamma]_{\phi(p)} \end{aligned}$$

$$\begin{aligned} T_p^*M &\longleftarrow T_{\phi(p)}N : T_p^*\phi \\ T_p^*\phi \cdot \alpha_{\phi(p)} &\longleftarrow\!\!\dashv \alpha_{\phi(p)} \end{aligned}$$

and are shown in Fig. 10.19. The pull-back of a one-form $\alpha_{\phi(p)} \in T_{\phi(p)}N$ is $T_p^*\phi \cdot \alpha_{\phi(p)} \in T_p^*M$ which therefore eats vectors $v_p \in T_pM$. Thus we defined the pull-back of a one-form by

$$\Big(T_p^*\phi \cdot \alpha_{\phi(p)}\Big)\big(v_p\big) \equiv \alpha_{\phi(p)}\Big(T_p\phi \cdot v_p\Big).$$

Also recall how the point that indexes the pull-back mapping is the point p from the range, not the domain. This is a rather unusual notational convention, but given the above definition it makes the connection between the pull-back and the push-forward maps obvious. Using the equivalence class of curves definition of vectors that we introduced in this section we have the pull-back given by

$$\boxed{\begin{array}{ll}\text{Pull-back of one-form } \alpha_{\phi(p)} \text{ by } T_p^*\phi & \begin{aligned}\Big(T_p^*\phi \cdot \alpha_{\phi(p)}\Big)\big([\gamma]_p\big) &\equiv \alpha_{\phi(p)}\Big(T_p\phi \cdot [\gamma]_p\Big) \\ &= \alpha_{\phi(p)}\Big(\cdot [\phi \circ \gamma]_{\phi(p)}\Big)\end{aligned}\end{array}}$$

This equation actually looks really simple in more traditional notation. Here we write the pull-back without the base point so you can see how simple it really is,

$$\boxed{\phi^*\alpha\big([\gamma]\big) = \alpha\big([\phi \circ \gamma]\big).}$$

Similarly, for a k-form $\omega_{\phi(p)} \in \bigwedge^k_{\phi(p)}(N)$ we get the pull-back mapping

$$\begin{aligned}\bigwedge\nolimits^k_p(M) &\longleftarrow \bigwedge\nolimits^k_{\phi(p)}(N) : T_p^*\phi \\ T_p^*\phi \cdot \omega_{\phi(p)} &\longleftarrow\!\shortmid\ \omega_{\phi(p)},\end{aligned}$$

which is defined as

$$\Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big(v_{p_1}, v_{p_2}, \ldots, v_{p_k}\big) \equiv \omega_{\phi(p)}\Big(T_p\phi \cdot v_{p_1}, T_p\phi \cdot v_{p_2}, \ldots, T_p\phi \cdot v_{p_k}\Big).$$

In terms of the equivalence class of curves definition of vectors we have the pull-back of a p-form ω given by

$$\boxed{\begin{array}{ll}\text{Pull-back of } p\text{-form } \omega_{\phi(p)} \text{ by } T_p^*\phi & \begin{aligned}\Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big([\gamma_1]_p, [\gamma_2]_p, \ldots, [\gamma_k]_p\big) &\equiv \omega_{\phi(p)}\Big(T_p\phi \cdot [\gamma_1]_p, T_p\phi \cdot [\gamma_2]_p, \ldots, T_p\phi \cdot [\gamma_k]_p\Big) \\ &= \omega_{\phi(p)}\Big([\phi \circ \gamma_1]_{\phi(p)}, [\phi \circ \gamma_2]_{\phi(p)}, \ldots, [\phi \circ \gamma_k]_{\phi(p)}\Big).\end{aligned}\end{array}}$$

We also give the definition of the pull-back of a zero-form. Recall that a zero-form on the manifold N is just a real-valued function $f : N \to \mathbb{R}$. The pull-back of a function is exactly what you would expect it to be,

$$\boxed{\begin{array}{ll}\text{Pull-back of zero-form } f \text{ by } T_p^*\phi & T_p^*\phi \cdot f \equiv \big(f \circ \phi\big)(p) = f\big(\phi(p)\big).\end{array}}$$

This definition looks very simple in traditional notation and without the base point included,

$$\boxed{\begin{array}{ll}\text{Pull-back of zero-form } f \text{ by } \phi^* & \phi^* \cdot f \equiv f \circ \phi.\end{array}}$$

In fact, we have already seen and used this in Sect. 6.7 in the proof of the third identity $\phi^* d = d\phi^*$.

10.4 Calculus on Manifolds

There is no reason we couldn't classify all the calculus you learned in your calculus and vector calculus classes as "calculus on a manifold." In these cases the manifold would simply be the Euclidian spaces $\mathbb{R}^2$ or $\mathbb{R}^3$. However, when one hears the phrase *calculus on manifolds* or perhaps alternative phrases like, *integration on manifolds* or even *analysis on manifolds* then the general assumption is that the manifold is more complicated than Euclidian space, that it is the sort of manifold we have introduced in this chapter, one with an atlas made up of coordinate charts (U_i, ϕ_i).

10.4.1 Differentiation on Manifolds

The word calculus of course implies both differentiation and integration. In this section we will make a few big-picture comments about differentiation on manifolds. The next subsection, as well as the next chapter, will focus on integration on manifolds. There are **three distinct notions** of differentiation on manifolds that one often encounters. All three of these notions work a little bit differently.

1. Exterior derivative: This is the kind of differentiation we are studying in this book and have already looked extensively at exterior differentiation in Chap. 4. Exterior differentiation acts on differential forms. When the differential form is a zero-form, or in other words simply a function, exterior differentiation is identical to the directional derivative from calculus. As we saw in Sect. 4.5, in a very real sense exterior differentiation is a generalization of the idea of directional derivatives from our general calculus classes. It is also this version of differentiation that gives us the generalized Stokes' theorem that is proved in Chap. 11. This generalizes Stokes' Theorem is an important component of integration on manifolds. Finally, in Sect. 4.3 we showed the local existence and uniqueness of the exterior differential operator. Now that we have the necessary concepts we will show global existence and uniqueness of the exterior differential operator below.
2. Lie derivative: Lie derivatives act on objects called tensors. Differential forms are in fact a subset of tensors and so it also makes perfect sense to refer to the Lie derivative of a differential form. Also, Lie differentiation has a very nice geometric meaning that is easier to visualize and understand than than exterior differentiation. We will introduce tensors in Appendix A and discuss the Lie derivative in Sect. A.7. There we derive several formulas that relate exterior differentiation with Lie differentiation. In fact, one of these formulas is used to prove, by induction, the global exterior derivative formula that was discussed from a somewhat geometric standpoint in Sect. 4.4.2.
3. Covariant derivative: Covariant derivatives will not be discussed at all in this book, but they rely on the concept of a connection that was discussed very briefly at the start of Sect. 10.1. As was mentioned, connections are additional structures on a manifold that "connect" tangent spaces in order to tell us what the parallel transport of a vector from one tangent space to another is. The essential idea is that knowing what constitute parallel vectors in different tangent spaces allows us to measure how tensors and forms change as their input vectors are parallel transported. Covariant derivatives are a second way to generalize directional derivatives from our general calculus classes.

You should recognize that all three of these notions of derivative are equally valid when discussing $\mathbb{R}^2$ or $\mathbb{R}^3$. It is just that in your introductory calculus classes you were only really interested in functions and not differential forms or tensors. On top of that there was a very natural, and intuitively understandable, Euclidian connection on $\mathbb{R}^2$ or $\mathbb{R}^3$ built into our manifold. Because of these considerations there was no need to actually ever discuss the various concepts of differentiation in a general calculus class, old-fashioned directional derivatives were enough. By the time you got to vector calculus, you actually did study exterior differentiation, but as you leaned in Chap. 9 exterior differentiation was given the special names gradient, curl, and divergence. In other words, the nature of exterior differentiation was masked by a different vocabulary and an attempt to avoid introducing forms. That is why everything was kept as a vector.

But in the more general and abstract setting we need to pay attention to these different concepts of differentiation. There are a large number of identities that relate these different ideas of derivative to each other. All of this makes the topic of calculus on a manifold quite a bit more complicated than calculus on the Euclidian spaces $\mathbb{R}^2$ or $\mathbb{R}^3$. Calculus on a manifold is really starting to get into the field of differential geometry, and all of the above mentioned ideas would really constitute the introductory topics for a first course in differential geometry.

With these comments made, we now turn our attention to some of the basics that are important in the context of differential forms. We begin by returning briefly to where we left off in Sect. 4.3. In that section we took four properties that we wanted exterior differentiation to have, called them axioms, and then derived a general formula for exterior differentiation. The fourth axiom basically stated that in the case of a zero-form that exterior differentiation had to be the same as the directional derivative, thereby ensuring that exterior differentiation was in fact a generalization of the directional derivative. We then used these axioms to derive a general formula for the exterior derivative. Once we had done that and had a single formula for the exterior derivative of a general differential form we knew that the exterior derivative operator d had to both exist and be unique.

Take a moment to review Sect. 4.3 carefully. Notice two things. First, that axiom four stated that in local coordinates, for each function f, $df = \sum \frac{\partial f}{\partial x_i} dx_i$. Second, we had supposed that we had the n-form

$$\alpha = \sum \alpha_{i_1 \cdots i_n} \, dx_{i_1} \wedge \cdots \wedge dx_{i_n}.$$

In both of these cases we were implicitly relying on a coordinate system, and in the context of that section we were implicitly relying on the Cartesian coordinate system. But a general manifold does not have a single coordinate system, instead it has an atlas $\{(U_i, \phi_i)\}$ of coordinate patches (U_i, ϕ_i). What was done in Sect. 4.3 only applies to a single coordinate patch (U_i, ϕ_i). Of course the same argument and computation can be done on each of the many coordinate patches, but all that means is that on each of the many coordinate patches we found a formula that gives the exterior derivative of a differential form, as long as that form is written in the coordinates of that particular coordinate patch.

Now we show that the exterior derivative operator d exists and is unique globally. That means that d exists and is unique over the whole manifold $M = \bigcup U_i$. On the coordinate patch (U_i, ϕ_i) we showed that the exterior derivative operator, which we will now label d_{U_i}, exists and is unique. We define the global differential operator d as $d = d_{U_i}$ on this coordinate patch. Similarly, on the coordinate patch (U_j, ϕ_j) we showed that the exterior derivative operator, which we will now label d_{U_j}, exists and is unique. Again, we define the global differential operator d as $d = d_{U_j}$ on this coordinate patch. The question is, what happens if $U_i \cap U_j \neq \varnothing$? Since d_{U_i} exists and is unique on U_i and d_{U_j} exists and is unique on U_j, then on $U_i \cap U_j$ we must have $d_{U_i} = d_{U_j} = d$. Since this is true on the intersection of all coordinate patches then we have d existing and unique globally, that is, over $M = \bigcup U_i$.

10.4.2 Integration on Manifolds

Now we turn our attention to integration on a manifold. Since there are several different ideas for differentiation on a manifold you should not be surprised that there are also different ideas of integration on a manifold. Here we will restrict ourselves to discussing integration of differential forms on a manifold.

In order for everything to turn out nicely we will assume the manifold is **oriented**. There are ways to do integration when a manifold is not oriented, but we will not discuss them here. An oriented n-dimensional manifold M is a manifold that has an n-form that is not zero at any point of the manifold. In other words, if $f dx_1 \wedge \cdots \wedge dx_n$ is an n-form on M then there is no $p \in M$ such that $f(p) = 0$. That implies that for all points $p \in M$ either $f(p) > 0$ or $f(p) < 0$. A form that has this property is called a **nowhere-zero n-form**. Thus the set of all nowhere-zero n-forms splits into two equivalence classes, one equivalence class that consists of all the everywhere positive nowhere-zero n-forms and an a second equivalence class that consists of all the everywhere negative nowhere-zero n-forms. By choosing one of these two equivalence classes we are specifying what is called an **orientation** of the manifold. We can think of an orientation as allowing us to find volumes on the manifold in a consistent way. Alternatively, we can think of an orientation as allowing us to orient all tangent spaces in a consistent way. The standard orientation of the manifold $\mathbb{R}^n$ is the equivalence class that contains the nowhere-zero n-form $dx_1 \wedge \cdots \wedge dx_n$. This orientation is often described as the orientation induced by $dx_1 \wedge \cdots \wedge dx_n$.

But notice what we did here. Saying a manifold M has an orientation induced by the n-form $dx_1 \wedge \cdots \wedge dx_n$ is all well and good as long as we have a manifold whose atlas consists of a single chart. But what if we do not? Consider Fig. 10.20 where we show a manifold with two charts, (U_i, ϕ_i) and (U_j, ϕ_j), where $U_i \cap U_j \neq \varnothing$. What would a volume form on U_i look like? Let us consider the standard volume form $dx_1 \wedge \cdots \wedge dx_n$, which is defined on all of $\mathbb{R}^n$ and hence is clearly defined on $\phi_i(U_i) \subset \mathbb{R}^n$. We can pull-back this standard volume form onto $U_i \subset M$. Thus $T^*\phi_i \cdot \big(dx_1 \wedge \cdots \wedge dx_n\big)$ is, in a sense, a "standard" volume form on $U_i \subset M$.

Similarly, what would a volume form on U_j look like? Like before we will consider the standard volume form on $\phi_j(U_i) \subset \mathbb{R}^n$, which we will call $dy_1 \wedge \cdots \wedge dy_n$ so as not to confuse it with $dx_1 \wedge \cdots \wedge dx_n$. We can pull-back this standard volume form onto $U_j \subset M$ to get the pull-back $T^*\phi_j \cdot \big(dy_1 \wedge \cdots \wedge dy_n\big)$ as a "standard" volume form on $U_j \subset M$. How do these two volume forms relate to each other on $U_i \cap U_j$ where they are both defined? In general they will not be the same. But remember what we are actually after. What we really want is to define a consistent volume form ω on all of M so that we can use it to define an orientation on M.

What does finding a "consistent volume form" ω on all of M actually mean? Consider Fig. 10.21. Suppose we have the standard volume form $dy_1 \wedge \cdots \wedge dy_n$ on $\phi_j(U_i)$. If there is a volume form $f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n$ on $\phi_i(U_i)$ such that on $U_i \cap U_j$ we have

$$T^*\phi_i\Big(f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n\Big) = T^*\phi_j\Big(dy_1 \wedge \cdots \wedge dy_n\Big)$$

then we could use these two volume forms, one on U_j and one on U_i to "stitch together" a volume form that is defined on $U_i \cup U_j$. By doing this repeatedly we can "stitch together" a volume form ω that is defined on all of M. But first we need to find the volume form $f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n$ on $\phi_i(U_i)$. Looking at Fig. 10.21 it is clear that the equality we want,

$$T^*\phi_i\Big(f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n\Big) = T^*\phi_j\Big(dy_1 \wedge \cdots \wedge dy_n\Big),$$

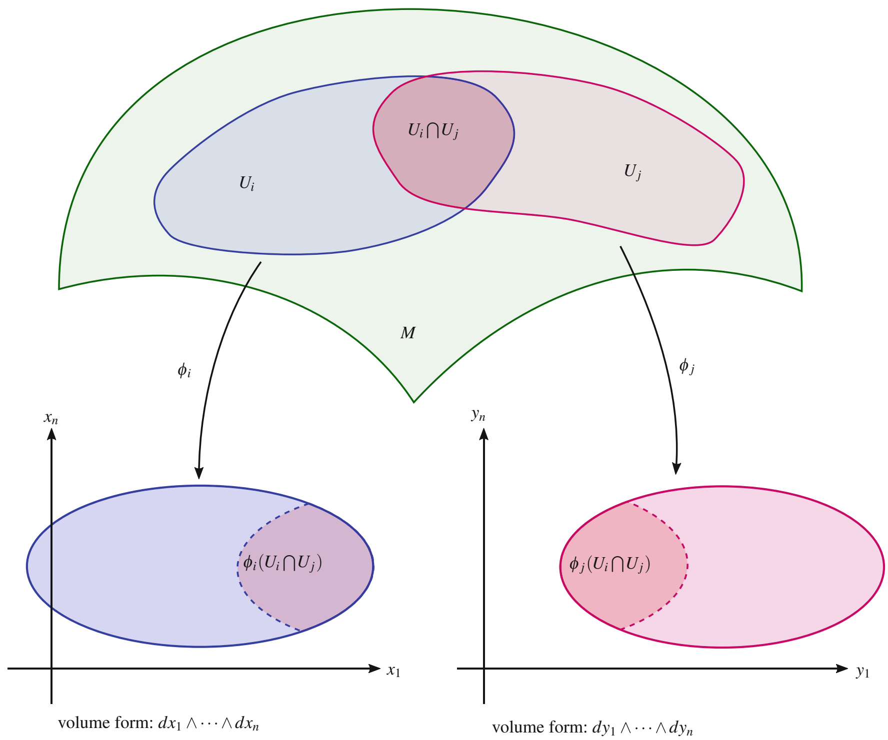

Fig. 10.20 The manifold M with two charts (U_i, ϕ_i) and (U_j, ϕ_j), where $U_i \cap U_j \neq \varnothing$. The standard volume for $dx_1 \wedge \cdots \wedge dx_n$ for $\mathbb{R}^n$ exists on both $\phi_i(U_i)$ and on $\phi_j(U_j)$

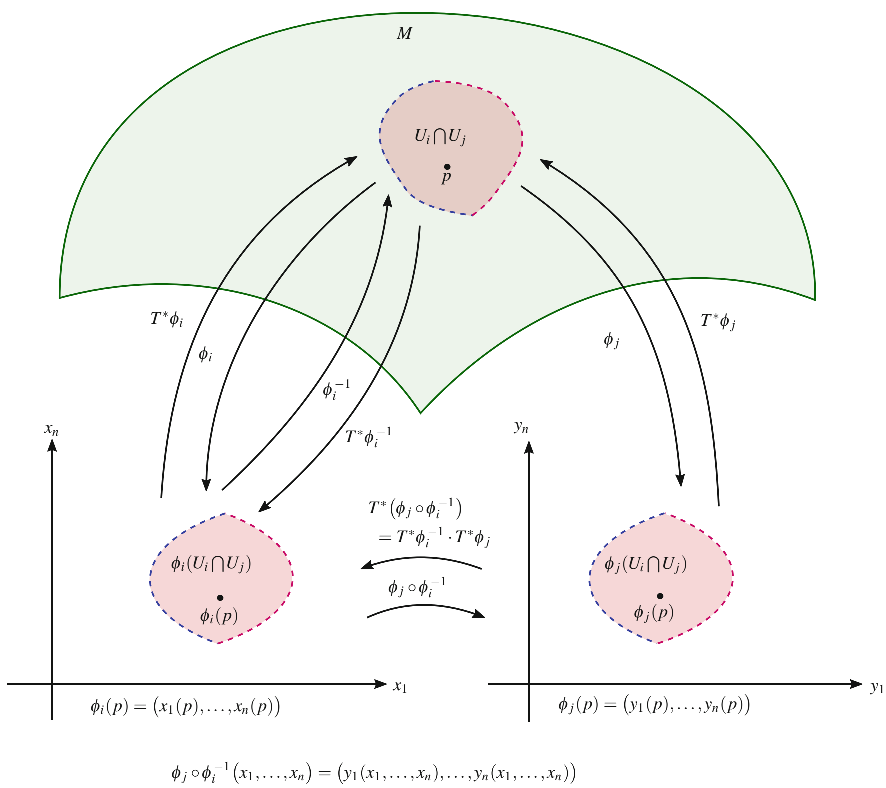

Fig. 10.21 Here we focus on $U_i \cap U_j$ from Fig. 10.20. All the necessary mappings to find $T^*\left(\phi_j \circ \phi_i^{-1}\right)\left(dy_1 \wedge \cdots \wedge dy_n\right)$

is equivalent to

$$\begin{aligned} f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n &= T^*\phi_i^{-1} \cdot T^*\phi_j \Big(dy_1 \wedge \cdots \wedge dy_n \Big) \\ &= T^*\Big(\phi_j \circ \phi_i^{-1} \Big)\Big(dy_1 \wedge \cdots \wedge dy_n \Big). \end{aligned}$$

The way we actually go about finding this is to rely on the mapping

$$\phi_j \circ \phi_i^{-1}(x_1, \ldots, x_n) = \Big(y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n) \Big)$$

to write the volume form $dy_1 \wedge \cdots \wedge dy_n$ in terms of $dx_1, \ldots, dx_n$,

$$\begin{aligned} dy_1 \wedge \cdots \wedge dy_n &= dy_1(x_1, \ldots, x_n) \wedge \cdots \wedge dy_n(x_1, \ldots, x_n) \\ &= \left(\frac{\partial y_1}{\partial x_1} dx_1 + \cdots + \frac{\partial y_1}{\partial x_n} dx_n \right) \wedge \cdots \wedge \left(\frac{\partial y_n}{\partial x_1} dx_1 + \cdots + \frac{\partial y_n}{\partial x_n} dx_n \right) \\ &= \underbrace{\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \end{vmatrix}}_{\substack{\text{determinant of} \\ \text{Jacobian of } \phi_j \circ \phi_i^{-1}}} dx_1 \wedge \cdots \wedge dx_n . \end{aligned}$$

Thus, we have

$$T^*\Big(\phi_j \circ \phi_i^{-1} \Big)\Big(dy_1 \wedge \cdots \wedge dy_n \Big) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \end{vmatrix} dx_1 \wedge \cdots \wedge dx_n .$$

In general the determinant of the Jacobian matrix of $\phi_j \circ \phi_j^{-1}$ could either be positive or negative, and usually we would not care. The pull-back of $dy_1 \wedge \cdots \wedge dy_n$ simply is what it is. However, right now we are interested in a consistent volume form ω on all of M which means, in $U_i \cap U_j$, we have

$$T^*\phi_i \Big(f(x_1, \ldots, x_n) dx_1 \wedge \cdots \wedge dx_n \Big) = T^*\phi_j \Big(dy_1 \wedge \cdots \wedge dy_n \Big).$$

Since $dx_1 \wedge \cdots \wedge dx_n$ and $dy_1 \wedge \cdots \wedge dy_n$ have the same orientation, in order for the right and the left sides of the above equation to have same orientation we would need to have the Jacobian of $\phi_j \circ \phi_j$ to be positive. Thus, as long as the mappings ϕ_i and ϕ_j are such that the transition function $\phi_j \circ \phi_i^{-1}$ has a Jacobian with a positive determinant, we can define a consistent volume form on $U_i \cup U_j$.

Extending this to all of M, we can find a consistent volume form on M as long as M has an atlas $\big\{(U_i, \phi_i)\big\}$ such that every transition function $\phi_j \circ \phi_i^{-1}$ has a Jacobian with a positive determinant. Another way to phrase this would be to say that a manifold M is orientable as long as it **admits** an atlas whose transition functions have a positive Jacobian determinant.

Now we turn to the next idea that we need to handle integration on manifolds. Suppose we have a manifold M with an atlas $\big\{(U_i, \phi_i)\big\}$. Since $\big\{(U_i, \phi_i)\big\}$ is an atlas of M we have $M = \bigcup U_i$. The set of coordinate patches $\big\{U_i\big\}$ is called a **covering** of M. For each coordinate patch U_i suppose we had a real-valued function $\varphi_i : M \to \mathbb{R}$ that was non-zero only for some subset V of U_i. In other words, if $p \notin V \subset U_i$ then $\varphi_i(p) = 0$. The function φ_i is said to be **subordinate** to the coordinate patch U_i. A *partition of unity* for M subordinate to the covering $\{U_i\}$ is a collection of of functions $\varphi_i : M \to \mathbb{R}$ such that

(a) for each coordinate patch U_i in the cover of M we have a function φ_i, which is subordinate to the coordinate patch U_i,
(b) such that $0 \leq \varphi_i(p) \leq 1$ for all $p \in U_i$, and
(c) for every $p \in M$ we have $\sum_i \varphi_i(p) = 1$.

If the manifold M is orientable it is generally assumed that the atlas used to make up the covering of M used in finding the partition of unity is in fact the atlas whose transition functions have a positive Jacobian determinant, that is, the atlas used to define the volume form ω on M.

The partition of unity allows us to "break up" any differential form α into a set collection of differential forms which are subordinate to the covering $\{U_i\}$,

$$\alpha = \sum_i \varphi_i \alpha.$$

It is clear that $\varphi_i \alpha$ is only non-zero on the coordinate patch U_i. We can only actually integrate α over regions that cover multiple coordinate patches by breaking it up into forms that are integrable on each coordinate patch. Thus for example, if we wanted to integrate α over M we would have

$$\begin{aligned}\int_M \alpha &= \int_M \sum_i \varphi_i \alpha \\ &= \sum_i \int_M \varphi_i \alpha \\ &= \sum_i \int_{U_i} \varphi_i \alpha\end{aligned}$$

where the last equality follows since $M = \bigcup_i U_i$ and $\varphi_i \alpha = 0$ outside of U_i. If we were integrating over a region $R \subset M$ then we have

$$\int_R \alpha = \sum_i \int_{U_i \cap R} \varphi_i \alpha.$$

So, being able to integrate a form on any region of M, or even on all of M, boils down to being able to integrate on a single coordinate patch. So now we will turn our attention to integrating a form α on a single coordinate patch. Notice, we change notation and simply use α instead of $\varphi_i \alpha$.

We begin by assuming α is an n-form on the n-dimensional manifold M and we want to find $\int_{U_i} \alpha$. All we have to do is pull-back α to $\phi_i(U_i)$ and integrate there, in other words,

$$\int_{U_i} \alpha = \int_{\phi_i(U_i)} T^* \phi_i^{-1} \cdot \alpha.$$

If you were asked to integrate the real-valued function $f : M \to \mathbb{R}$ then what this really means is that we want to integrate the n-form $f\omega$, where ω is the volume form on M. Thus we use

$$\begin{aligned}\int_{U_i} f\omega &= \int_{\phi_i(U_i)} T^* \phi_i^{-1} \cdot (f\omega) \\ &= \int_{\phi_i(U_i)} f \circ \phi_i^{-1}\, T^* \phi_i^{-1} \cdot \omega.\end{aligned}$$

Question 10.4 Show the second equality in the above equation.

In order to integrate a k-form β on an n-dimensional manifold M, where $k < n$, then we need need a k-dimensional submanifold to integrate over. Without getting into the technical definition of what a submanifold is, you can think of it essentially as a parameterized surface $\Sigma \subset M$ that is given by a mapping $\Phi : \mathbb{R}^k \to M$. This is exactly what was done in Sect. 7.4 where we integrated over one- and two-dimensional parameterized surfaces. We can then integrate β on the submanifold Σ by

$$\int_{\Sigma} \beta = \int_{\Phi^{-1}(\Sigma)} T^*\Phi \cdot \beta.$$

So, even though the discussion in this section has been quite abstract and theoretical, it turns out that we have already done a number of concrete examples for what we are discussing.

This actually covers only the very basics of integration on manifolds. In Chap. 11 we will expand upon this topic by proving a very important theorem regarding integration on manifolds, a theorem that is central to a great deal of mathematics, called the generalized Stokes' theorem, or more usually, simply called Stokes' theorem.

10.5 Summary, References, and Problems

10.5.1 Summary

An n-dimensional manifold is a space M that can be completely covered by a collection of local coordinate neighborhoods U_i with one-to-one mappings $\phi_i : U_i \to \mathbb{R}^n$, which are called a coordinate maps. Together U_i and ϕ_i are called a coordinate patch or a chart, which is generally denoted as (U_i, ϕ_i). The set of all the charts together, $\big\{(U_i, \phi_i)\big\}$, is called a coordinate system or an atlas of M. Since the U_i cover all of M we write that $M = \bigcup U_i$. Also, since ϕ_i is one-to-one it is invertible, so ϕ_i^{-1} exists and is well defined. If two charts have a non-empty intersection, $U_i \cap U_j \neq \varnothing$, then the functions $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ are called transition functions.

$$\begin{aligned}
U_i \quad &: \quad \text{coordinate neighborhood} \\
\phi_i : U_i \to \mathbb{R}^n \quad &: \quad \text{coordinate map} \\
(U_i, \phi_i) \quad &: \quad \text{coordinate patch/chart} \\
\big\{(U_i, \phi_i)\big\} \quad &: \quad \text{coordinate system/atlas} \\
\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n \quad &: \quad \text{transition function}
\end{aligned}$$

Suppose $r \in U_i \cap U_i \neq \emptyset$, then $\phi_i(r) \in \mathbb{R}^n$ and $\phi_j(r) \in \mathbb{R}^n$. Furthermore, $\phi_j \circ \phi_i^{-1}$ sends $\phi_i(U_i \cap U_j) \subset \mathbb{R}^n$ to $\phi_j(U_i \cap U_j) \subset \mathbb{R}^n$. That is, $\phi_j \circ \phi_i^{-1}$ is a map of a subset of $\mathbb{R}^n$ to another subset of $\mathbb{R}^n$, and so the mapping $\phi_j \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ with domain $\phi_i(U_i \cap U_j)$ and range $\phi_j(U_i \cap U_j)$ is the sort of mapping that we know how to differentiate from multivariable calculus. A differentiable manifold is a set M, together with a collection of charts (U_i, ϕ_i), where $M = \bigcup U_i$, such that every mapping $\phi_j \circ \phi_i^{-1}$, where $U_i \cap U_j \neq \emptyset$, is differentiable.

We define the tangent space of M at point p, T_pM, in terms of curves on M. If two curves have the same range close to p and have the same parametrization close to p they are called equivalent, which is denoted $\sim$. The set of all equivalent curves is called an equivalence class and is defined by

$$[\gamma_1] \equiv \big\{\gamma \;\big|\; \gamma \sim \gamma_1\big\}.$$

Each equivalence class of curves at a point p is defined to be a tangent vector at p. The tangent space of M at p is defined as the set of all tangent vectors, that is, equivalence classes of curves, at the point p,

$$T_pM = \Big\{ [\gamma]_p \;\Big|\; \gamma : (-\epsilon, \epsilon) \to M \text{ and } \gamma(0) = p\Big\}.$$

The relation between the equivalence class (intrinsic) definition and the vector (extrinsic) definitions of the tangent space is summarized,

Equivalence Class Definition	$\Longleftrightarrow$	Vector Definition
$v_p = [\gamma]_p, \quad p = \gamma(0)$ $= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]$		$$\begin{aligned} v_p &= [\gamma']_p = [\gamma]'_p \\ &= \Big[\big(\gamma'_1(t), \ldots, \gamma'_n(t)\big)\Big]_p \\ &= \Big[\big(\gamma_1(t), \ldots, \gamma_n(t)\big)\Big]'_p \\ &= \begin{bmatrix} \gamma'_1(t=0) \\ \vdots \\ \gamma'_n(t=0) \end{bmatrix}_p \\ &= \begin{bmatrix} \left.\frac{\partial \gamma_1(t)}{\partial t}\right\|_{t=0} \\ \vdots \\ \left.\frac{\partial \gamma_n(t)}{\partial t}\right\|_{t=0} \end{bmatrix}_p . \end{aligned}$$

The equivalence class of curves definition gives us another way to think about the chain rule,

Chain Rule for curves
$$[f \circ \phi^{-1} \circ \phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi\circ\gamma(0)} \cdot [\phi \circ \gamma]'_0 = [f \circ \phi^{-1}]'_{\phi(p)} \cdot [\phi \circ \gamma]'_0.$$

For the equivalence class of curves definition of vectors the push-forward of $[\gamma]_p$ by $T_p\phi$ is

Push-foraward of vector $[\gamma]_p$ by $T_p\phi$
$$T_p\phi \cdot [\gamma]_p = [\phi \circ \gamma]_{\phi(p)}.$$

Notationally there are several different ways to write this, all of which represent the same thing,

Ways to write push-foraward of vector $[\gamma]_p$by $T_p\phi$
$$\begin{aligned} T_p\phi \cdot [\gamma]_p &\equiv \phi_*(p)\big([\gamma]_p\big) \\ &\equiv \phi_*(p) \cdot [\gamma]_p \\ &\equiv \phi_*\big([\gamma]_p\big) \\ &\equiv [\phi \circ \gamma]_{\phi(p)}. \end{aligned}$$

In the equivalence class of curves definition of vectors the pull-back of a p-form ω is given by

Pull-back of p-form $\omega_{\phi(p)}$ by $T_p^*\phi$
$$\begin{aligned} \Big(T_p^*\phi \cdot \omega_{\phi(p)}\Big)\big([\gamma_1]_p, [\gamma_2]_p, \ldots, [\gamma_k]_p\big) &\equiv \omega_{\phi(p)}\Big(T_p\phi \cdot [\gamma_1]_p, T_p\phi \cdot [\gamma_2]_p, \ldots, T_p\phi \cdot [\gamma_k]_p\Big) \\ &= \omega_{\phi(p)}\Big([\phi \circ \gamma_1]_{\phi(p)}, [\phi \circ \gamma_2]_{\phi(p)}, \ldots, [\phi \circ \gamma_k]_{\phi(p)}\Big) \end{aligned}$$

and the pull-back of a zero form (function) $f : N \to \mathbb{R}$ is

Pull-back of zero-form f by $T_p^*\phi$
$$T_p^*\phi \cdot f \equiv \big(f \circ \phi\big)(p) = f\big(\phi(p)\big).$$

This definition looks very simple in traditional notation and without the base point included,

Pull-back of zero-form f by ϕ^*
$$\phi^* \cdot f \equiv f \circ \phi.$$

While this has been far from a complete introduction to manifolds, we have attempted to provide the basic ideas associated with a general manifold M. In particular, the idea of charts and atlases are the essential ingredients in the definition of a general manifold. We have also tried to show, particularly in Sect. 10.4, how important being careful with regards to charts is and how it does add an extra layer of complexity to the study of calculus on manifolds. It is usually not difficult to deal with, but attention does need to be paid to this extra layer of complexity. Since this is primarily a book on differential forms we have tried to avoid this extra layer of complexity throughout most of the book by sticking to the Euclidian manifolds $\mathbb{R}^n$, which all have an atlas consisting of one chart, the Euclidian chart that we are very familiar with.

10.5.2 References and Further Reading

There are literally hundreds of books on manifold theory and everything presented here is quite standard. If anything we have sacrificed mathematical rigor for readability and comprehensibility given this introduction to manifolds is so short. This chapter is meant to be simply a very basic introduction to some of the fundamentals, enough to get the reader started in more advanced books. The following five references are all good and are given in order of increasing levels of difficulty; Munkres [35], Walschap [47], Renteln [37], Martin [33], and Conlon [10]. In addition, the somewhat olders book by Bishop and Crittenden [5] and do Carmo [14] are very nice, if one can get hold of them.

10.5.3 Problems

Question 10.5 Write the following vectors on the manifold $\mathbb{R}^2$ as an equivalence class of curves in three different ways. (It may be useful to consider using trigonometric or exponential functions.)

a) $\begin{bmatrix} 2 \\ 3 \end{bmatrix}_{(2,1)}$ *b)* $\begin{bmatrix} -4 \\ 5 \end{bmatrix}_{(3,-2)}$ *c)* $\begin{bmatrix} 7 \\ -3 \end{bmatrix}_{(-3,-1)}$ *d)* $\begin{bmatrix} -3 \\ -7 \end{bmatrix}_{(-4,8)}$ *e)* $\begin{bmatrix} 3 \\ 1 \end{bmatrix}_{(3,-2)}$

Question 10.6 Using the vectors v_p from Question 10.5 written as the equivalence class of a curve, find $v_p[f_i]$ for the below functions f_i. Verify that you obtain the same answer regardless of which of the three representations of the curve you use.

a) $f_1(x, y) = xy^2$ *b)* $f_2(x, y) = y + \cos(x)$ *c)* $f_3(x, y) = e^x + 3y - 4$
d) $f_4(x, y) = \sqrt{xy} + xy$ *e)* $f_5(x, y) = (x + 2y)^2$

Question 10.7 Write the following vectors on the manifold $\mathbb{R}^3$ as an equivalence class of curves in three different ways. (It may be useful to consider using trigonometric or exponential functions.)

a) $\begin{bmatrix} 4 \\ -2 \\ -8 \end{bmatrix}_{(5,-2,4)}$ *b)* $\begin{bmatrix} 3 \\ -6 \\ 5 \end{bmatrix}_{(3,-4,5)}$ *c)* $\begin{bmatrix} -4 \\ 7 \\ -1 \end{bmatrix}_{(2,4,-3)}$ *d)* $\begin{bmatrix} 5 \\ 9 \\ -9 \end{bmatrix}_{(6,-4,2)}$ *e)* $\begin{bmatrix} -5 \\ -2 \\ -3 \end{bmatrix}_{(7,3,-4)}$

Question 10.8 Using the vectors v_p from Question 10.7 written as the equivalence class of a curve, find $v_p[f_i]$ for the below functions f_i. Verify that you obtain the same answer regardless of which of the three representations of the curve you use.

a) $f_1(x, y, z) = xy^2z$ *b)* $f_2(x, y, z) = y + \cos(x) + \sin(z)$ *c)* $f_3(x, y, z) = e^x + ye^z$
d) $f_4(x, y, z) = \sqrt{xyz} + xyz$ *e)* $f_5(x, y, z) = (x + 2y - 3z)^2$

Question 10.9 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(2y, 3x\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.10 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(3x + 2y, 3x - 2y\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.11 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(xy, xy^3\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.5 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.12 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(-2z, 3x, -4y\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.13 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(3x + 2y + z, 3x - 2y + z, 5z + 7\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.14 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(xy, yz, zx\big)$. Using the equivalence class of curves version of vectors v_p in Question 10.7 find $T\phi \cdot v_p$. Verify your answer does not depend on which representative curve you use.

Question 10.15 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(2y, 3x\big)$ and let $\alpha = 3v\ du + 2u\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.16 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(3x + 2y, 3x - 2y\big)$ and let $\alpha = \cos(u)\ du + \sin(v)\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.17 Let $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\phi(x, y) = \big(u(x, y), v(x, y)\big) = \big(xy, xy^3\big)$ and let $\alpha = (2u+v)\ du + u^2v^3\ dv$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.5. Verify your answer does not depend on which representative curve you use.

Question 10.18 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(-2z, 3x, -4y\big)$ and let $\alpha = 2w\ du + 3v\ dv + 4u\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Question 10.19 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(3x + 2y + z, 3x - 2y + z, 5z + 7\big)$ and let $\alpha = \sin(v)\ du + \cos(u)\ dv + 3w\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Question 10.20 Let $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\phi(x, y, z) = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big) = \big(xy, yz, zx\big)$ and let $\alpha = (u + v)\ du + (v + w)\ dv + (w + u)\ dw$. Evaluate $T^*\phi \cdot \alpha$ at each of the vectors in Question 10.7. Verify your answer does not depend on which representative curve you use.

Chapter 11
Generalized Stokes' Theorem

The generalized version of Stokes' theorem, henceforth simply called Stokes' theorem, is an extraordinarily powerful and useful tool in mathematics. We have already encountered it in Sect. 9.5 where we found a common way of writing the fundamental theorem of line integrals, the vector calculus version of Stokes' theorem, and the divergence theorem as $\int_M d\alpha = \int_{\partial M} \alpha$. More precisely Stokes' theorem can be stated as follows.

Theorem 11.1 (Stokes' Theorem) *Let M be a smooth oriented n-dimensional manifold and let α be an $(n-1)$-form on M. Then*

$$\int_M d\alpha = \int_{\partial M} \alpha,$$

where ∂M is given the induced orientation.

Its proof appears in pretty much all the standard texts of differential geometry and is done with various levels of abstraction and rigor. As with much of this book we will try to strike a balance between understandability, rigor, and abstraction. In particular, the proof we give works for "nice" manifolds M. Hopefully at the end of this chapter you will understand the big picture well enough that you could follow a more mathematically rigorous version of the proof.

The general strategy that the proof follows is that it is proved for a very specific and straight-forward case, and we use the simple case to prove harder and harder cases. In essence we "bootstrap" our way up to the general version. Each section in this chapter, except the last section, is essentially one of the "bootstrap" steps. Understanding the general ideas is not so difficult, but getting a totally rigorous and airtight proof can sometimes feel like an exercise in minutia, trivialities, and nitpicking. The final section of the chapter uses the visualization techniques developed in Chap. 5 to visualize what Stokes' theorem is saying in three dimensions, at least insofar as these visualization techniques work.

Finally, proofs of Stokes' theorem are either based on k-dimensional unit cubes, which is the route we will follow, or sometimes on k-dimensional simplices, usually just called k-simplices. (This is one of those rare English words that has two plurals is use, so sometimes you will see the word simplexes as well.) Applications of Stoke's theorem to homology and cohomology make it useful to use k-simplicies. While it is possible to move between the unit k-cubes and k-simplices this does add an almost overwhelmingly tedious layer to the bootstrapping process, so we will not do that step here. In reality this step adds virtually nothing to understanding the big picture, though it is something you should be aware of and on the lookout for when you read other proofs of Stokes' theorem.

11.1 The Unit Cube I^k

The very first thing we need to do is discuss the orientations of unit k-cubes in $\mathbb{R}^n$, where $n \geq k$. Unit k-cubes are generally denoted by I^k. After looking at the orientations of k-cubes I^k we will then look at the boundary of I^k and the orientations of the boundary pieces. Once we have a firm handle on this we can then prove Stokes' theorem for the case of the unit k-cube I^k.

J. P. Fortney, *A Visual Introduction to Differential Forms and Calculus on Manifolds*,
https://doi.org/10.1007/978-3-319-96992-3_11

The 0-Cube I^0

First we consider the 0-cube. The 0-cube actually isn't very interesting, it is just a point. For convenience sake we will place our cube at the origin. So we can consider the 0-cube in $\mathbb{R}^n$ to simply be a point which we will denote as $\{0\}$. Being degenerate the 0-cube does not have an orientation in the same way that the other k-cubes do, though as we will see we can assign it either a positive or a negative orientation as necessary.

The 1-Cube I^1

Next we will look at the unit 1-cube $I^1 = \{x_1 \in \mathbb{R} \mid 0 \leq x_1 \leq 1\}$, which is Fig. 11.1. We will say that the orientation of $I^1 \subset \mathbb{R}^1$ is determined by the volume form dx_1. How does that work? Consider the unit vector $[1]$ based at the origin pointing in the direction of the 1-cube. We have $dx_1([1]) = 1$ which is positive, so this gives us the positive orientation on I^1. The opposite direction $-[1]$ gives $dx_1(-[1]) = -1$, which clearly gives us the negative orientation, see Fig. 11.2.

The 2-Cube I^2

Now let us consider the 2-cube $I^2 = \{(x_1, x_2) \mid 1 \leq x_i \leq 1,\ i = 1, 2\}$. Again, the volume form $dx_1 \wedge dx_2$ determines (often we say it induces) an orientation on I^2. Consider the two unit vectors

$$e^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } e^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The 2-cube is exactly the parallelepiped spanned by these two unit vectors as shown in Fig. 11.3. Using the two-dimensional volume form $dx_1 \wedge dx_2$ we have

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 > 0,$$

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1 < 0.$$

Fig. 11.1 The unit cube in one dimension shown on the x_1-axis

Fig. 11.2 The two orientations of the one-dimensional unit cube. The orientations are determined using the one-dimensional volume form dx_1 and are generally called positive (left) and negative (right)

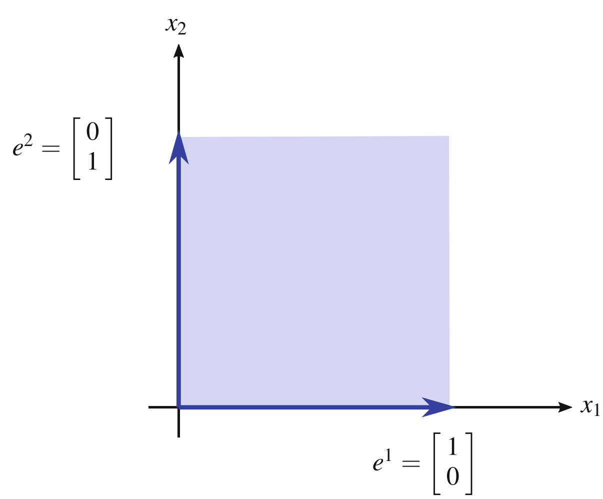

Fig. 11.3 The unit cube in two-dimensions is the parallelepiped spanned by the unit vectors e^1 and e^2

Thus the ordering of the unit vectors determines the orientation of the 2-cube. The ordering

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

denotes the positive orientation and the orientation

$$\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

denotes the negative orientation. The positive orientation orientation is usually shown graphically as in Fig. 11.4 while the negative orientation is usually show graphically as in Fig. 11.5.

The 3-Cube I^3

Now we will take a look at the 3-cube $I^3 = \{(x_1, x_2, x_3) \mid 0 \le x_i \le 1,\ i = 1, 2, 3\ \}$, shown in Fig. 11.6. The 3-cube is exactly the parallelepiped spanned by the vectors

$$e^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ e^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The volume form $dx_1 \wedge dx_2 \wedge dx_2$ determines, or induces, an orientation on I^3. Clearly we have

$$dx_1 \wedge dx_2 \wedge dx_3 \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 > 0.$$

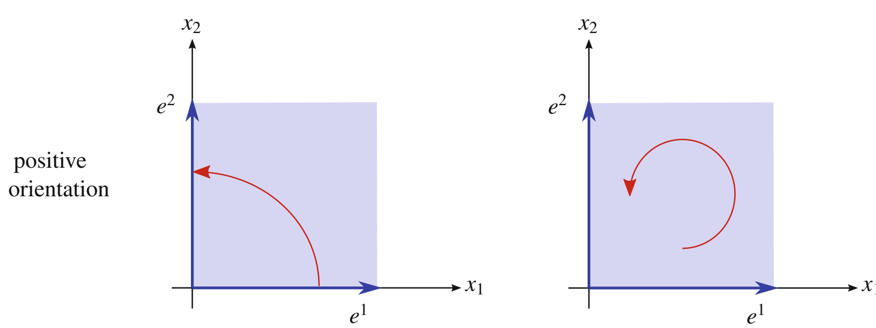

Fig. 11.4 Using the volume form $dx_1 \wedge dx_2$ we see that the ordering $\{e^1, e^2\}$, that is, e^1 followed by e^2, gives a positive orientation. Shown are two ways we can indicate a positive orientation

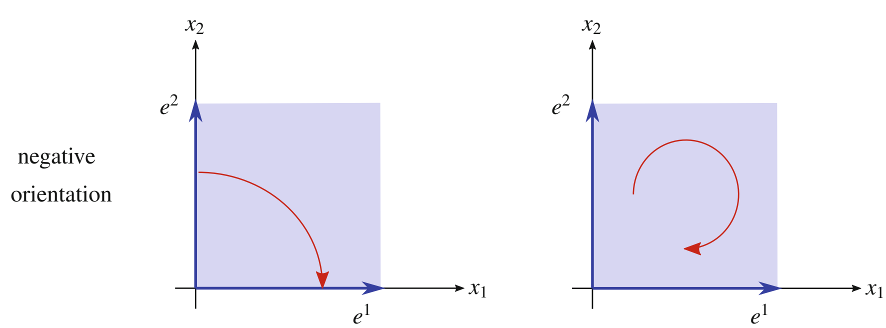

Fig. 11.5 Using the volume form $dx_1 \wedge dx_2$ we see that the ordering $\{e^2, e^1\}$, that is, e^2 followed by e^1, gives a negative orientation. Shown are two ways we can indicate a negative orientation

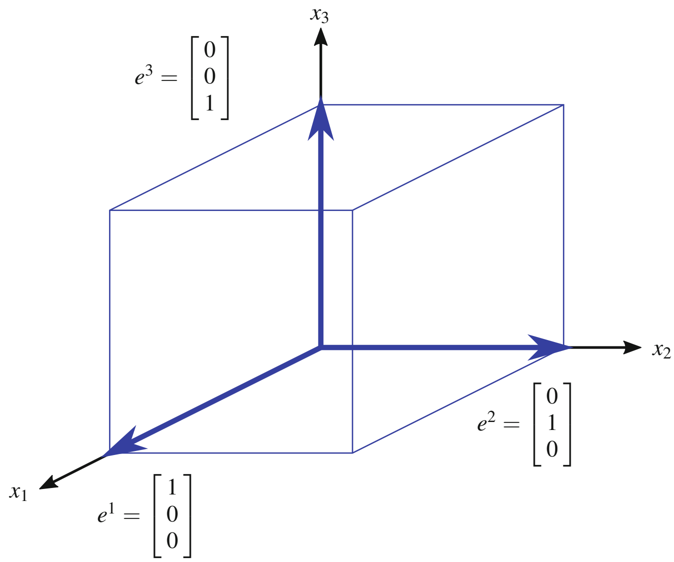

Fig. 11.6 The unit cube in three dimensions is the parallelepiped spanned by the three unit vectors e^1, e^2, and e^3

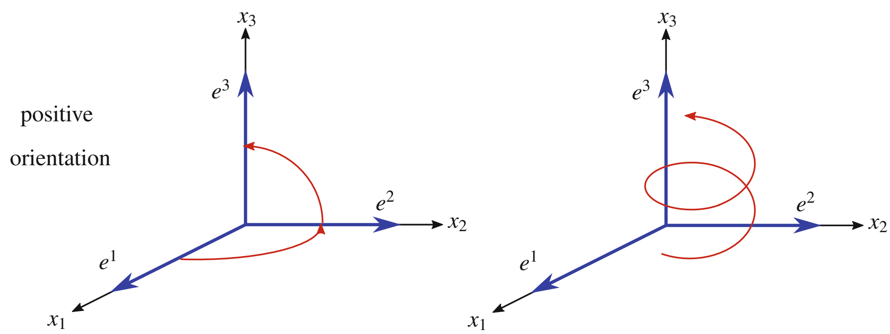

Fig. 11.7 Using the volume form $dx_1 \wedge dx_2 \wedge dx_3$ we see that the ordering $\{e^1, e^2, e^3\}$, that is, e^1 followed by e^2 followed by e^3, gives a positive orientation. Shown are two ways we can indicate a positive orientation. Positive orientation in three dimensions is generally called the "right-hand rule"

Thus the vector ordering

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \{e^1, e^2, e^3\}$$

is considered a positive orientation. The positive orientation is usually show graphically in Fig. 11.7. This is exactly the so called "right-hand rule." But notice, we also have $dx_1 \wedge dx_2 \wedge dx_3(e_2, e_3, e_1) = 1$ and so $\{e_2, e_3, e_1\}$ also is a positive orientation. Similarly, $dx_1 \wedge dx_2 \wedge dx_3(e_3, e_1, e_2) = 1$ so $\{e_3, e_1, e_2\}$ is yet another positive orientation. So we see that there are three different orderings that give a positive orientation of I^3. Next notice that $dx_1 \wedge dx_2 \wedge dx_3(e_2, e_1, e_3) = -1 < 0$ so $\{e^2, e^1, e^3\}$ gives us a negative orientation, which you could consider a "left-hand rule." See Fig. 11.8.

Question 11.1 What are the three orderings that give a negative orientation to I^3?

So we can see that choosing an order of our basis elements is not a very efficient way to label or determine k-cube orientations. That is why we use the standard volume form $dx_1 \wedge dx_2 \wedge dx_3$, that is the volume form with the indices in numeric order. Of course, this actually also boils down to choosing an ordering for our vector space basis elements, but we have to start somewhere. The standard volume form induces positive and negative orientations on each unit cube, depending on how we order the unit vectors that are used to generate the cube.

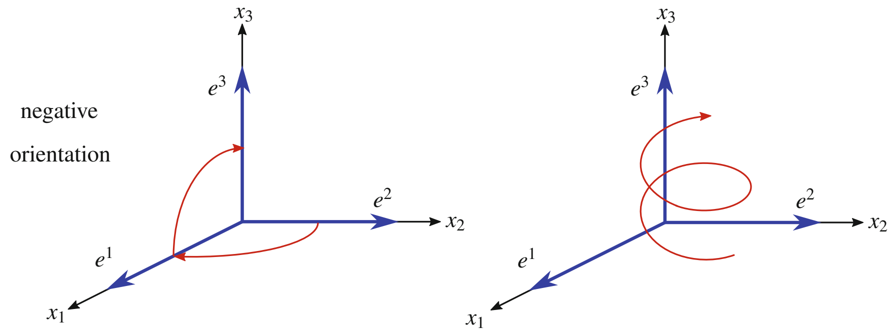

Fig. 11.8 Using the volume form $dx_1 \wedge dx_2 \wedge dx_3$ we see that the ordering $\{e^2, e^1, e^3\}$, that is, e^2 followed by e^1 followed by e^3, gives a negative orientation. Shown are two ways we can indicate a negative orientation. Negative orientation in three dimensions is generally called the "left-hand rule"

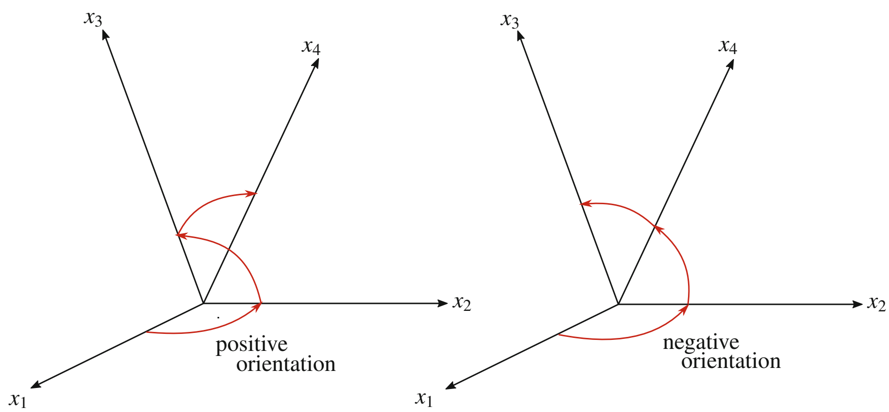

Fig. 11.9 An attempt to draw positive (left) and negative (right) orientations for a four dimensional space. Since we have four axes they can no longer be drawn as all perpendicular to the three others, but it should be understood that each of the depicted axis is in fact perpendicular to the remaining three

The k-Cube I^k

For I^k, where $k > 3$, it becomes impossible to draw either the k-cube or orientations, but the idea is intuitive enough. The k-cube is the parallelepiped spanned by the unit vectors $e^1, e^2, \dots, e^k$. And given a volume form $d_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ the volume form induces an orientation. For example, given the volume form $dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4$ any ordering of e_1, e_2, e_3, e_4 that results in a positive volume is called a positive orientation of I^4 and any ordering that results in a negative volume is called a negative orientation. We attempt to "draw" a positive and negative orientation for I^4 in Fig. 11.9.

Question 11.2 Find all the positive and negative orientations of e_1, e_2, e_3, e_4.

Boundaries of k-Cubes, ∂I^k

Now that we have some idea about the orientations of unit cubes we want to figure out what the boundaries of a unit cube are. Of course you have some intuitive idea of what boundaries are, but our goal is to make these ideas mathematically precise. We will also consider the idea of the orientations of the boundaries, which may be a little confusing at first.

Since the 0-cube is just a point, it does not have a boundary. We did not originally consider the orientation of a lone point I^0, but in the context of the boundary of I^1 we will assign an orientation to a point. First we will give the general definition of boundary of a k-cube, and then we will see some concrete examples of boundaries.

Given a k-cube I^k

$$I^k = \left\{ (x_1, \dots, x_k) \;\middle|\; 0 \le x_i \le 1,\ i = 1, \dots, k \right\}$$

we define the $(i, 0)$-face by letting $x_i = 0$, so we have

$$I^k_{(i,0)} = I^k(x_1, \ldots, x_{i-1}, \underbrace{0}_{x_i=0}, x_{i+1}, \ldots, x_k)$$
$$= \Big\{ (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) \mid (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) \in I^k \Big\}$$

and the $(i, 1)$-face by letting $x_i = 1$, so we have

$$I^k_{(i,1)} = I^k(x_1, \ldots, x_{i-1}, \underbrace{1}_{x_i=1}, x_{i+1}, \ldots, x_k)$$
$$= \Big\{ (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k) \mid (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_k) \in I^k \Big\}.$$

Each face $I^k_{(i,a)}$, where $a = 0, 1$, will be given the orientation determined by $(-1)^{i+a}$. If $(-1)^{i+a} = 1$ the orientation is positive and if $(-1)^{i+a} = -1$ then the orientation is negative. We will see, through the use of examples, why this is the appropriate way to define the face orientation. The boundary of I^k, denoted as ∂I^k, is the collection of all the faces of I^k along with the determined orientation. Abstractly we will write

$$\partial I^k = \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} I^k_{(i,a)},$$

though keep in mind that we are of course not adding these faces in the usual sense. This is simply our way of denoting the collection of faces, along with their respective orientations, that make up the boundary of I^k.

Finding ∂I^1
Let us see how this works for I^1. We have

$$I^1_{(1,0)} = \big\{(0)\big\},$$
$$I^1_{(1,1)} = \big\{(1)\big\}.$$

Next we see that $I^1_{(1,0)}$ has a negative orientation since $(-1)^{1+0} = -1$ and $I^1_{(1,1)}$ has a positive orientation since $(-1)^{1+1} = 1$. This is shown in Fig. 11.10. Finally, we would denote the boundary of I^1 by

$$\partial I^k = \sum_{i=1}^{1} \sum_{a=0}^{1} (-1)^{i+a} I^i_{(i,a)}$$
$$= I^1_{(1,1)} - I^1_{(1,0)}.$$

Finding ∂I^2
Next we turn our attention to I^2. The 2-cube has four faces given by

$$I^2_{(1,0)} = \Big\{ (0, x_2) \mid (0, x_2) \in I^2 \Big\},$$
$$I^2_{(1,1)} = \Big\{ (1, x_2) \mid (1, x_2) \in I^2 \Big\},$$

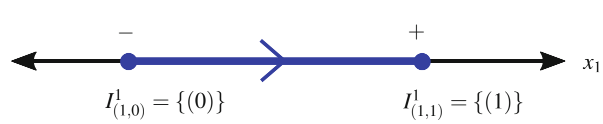

Fig. 11.10 The positively oriented one-cube I^1 along with the faces $I^1_{(1,0)}$ with a negative orientation, and the face $I^1_{(1,1)}$ with a positive orientation