MANIFOLD NOTES PART I
(DRAFT)

LARRY SUSANKA

Contents

Date: October 8, 2020.

The path through the following material has been guided by a number of sources, including [1] Michael Spivak's compendium "A Comprehensive Introduction to Differential Geometry Volume 1" and [2] John Lee's "Introduction to Smooth Manifolds 2nd Ed." and [3] Boothby's "An Introduction to Differentiable Manifolds and Riemannian Geometry." We assume a number of facts from Linear Algebra, Multi-Variable Calculus and Topology. Part I discusses the general set-up and parts of Differential Geometry that can be dealt with "point-by-point." Using first derivatives and linearization we attach a vector space to each point on a manifold and build structures corresponding to Linear Algebra. Ideas involving curvature and rates-of-change of these entities, second derivative concepts, will follow in Part II.

1. Topological Manifolds

In this note a **topological manifold** $\mathcal{M}$ will be a topological space with four "regularity" properties. In order to avoid pathologies not obviously relevant to our immediate purpose we will assume that $\mathcal{M}$ is **Hausdorff and second countable.**

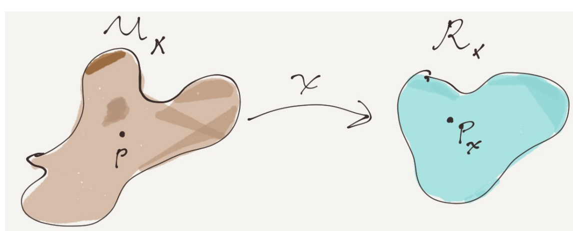
The third and most important property is that $\mathcal{M}$ should be **locally Euclidean.** That means some neighborhood of each point of $\mathcal{M}$ should be homeomorphic to an open subset of $\mathbb{R}^n$ with usual topology from the Euclidean metric.

Specifically, for each point in the manifold there is an open set containing that point and an open subset of $\mathbb{R}^n$ and a homeomorphism—a continuous function with continuous inverse—between these two open sets. Every point in $\mathcal{M}$ has a neighborhood that "looks like" $\mathbb{R}^n$, at least so far as can be determined by topological considerations.

We will try to be consistent about notation throughout this work, and there is (unfortunately) quite a bit of it to absorb and remember.

For instance we are going to be working with functions $H\colon A \to B$ and conceive of H as a vehicle to represent members of A inside B. When it reduces visual clutter in chains of compositions and to emphasize that $H(p)$ is "really is just p transported by H to a new environment" we will often denote $H(p)$ by $\boldsymbol{p_H}$.

As an example, we will use multiple homeomorphisms of the type $\boldsymbol{x\colon U_x \to R_x}$ mentioned above, where x has open domain $\boldsymbol{U_x} \subset \mathcal{M}$ and open range $\boldsymbol{R_x} \subset \mathbb{R}^n$. So if $p \in U_x$ we will often denote $x(p) \in \mathbb{R}^n$ by p_x.



The homeomorphism $x\colon U_x \to R_x$ is called a **coordinate map around p** whenever $p \in U_x$. The set U_x itself will be called a **coordinate patch.**

We will usually denote coordinate maps by a lower case letter such as

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} : U_x \to R_x.$$

The individual real-valued functions $x^i\colon U_x \to \mathbb{R}$ are called **coordinates** and the ordered n-tuple $x(p)$ is called, elliptically, **the coordinates of p with respect to the coordinate map x**.

Members of $\mathbb{R}^n$ are columns, and we will be scrupulous about displaying them in this way. For obvious typographical reasons some authors represent them on the page as rows but if they are consistent with this common matrix operations become awkward.

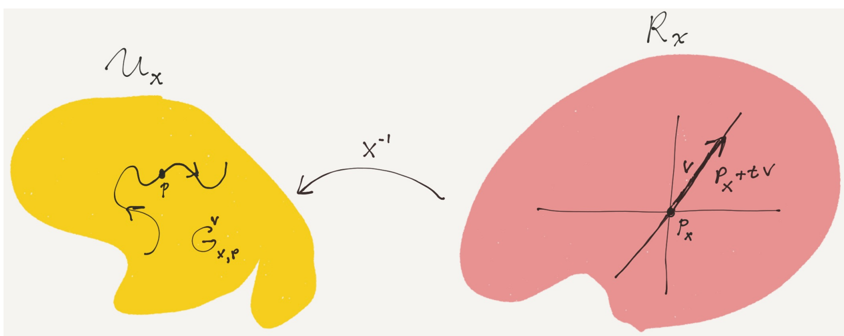
Columns must be distinguished from row matrices, the members of $\mathbb{R}^{n*}$. We will need both. Coordinates of row matrices may be separated by commas to avoid inadvertent concatenation in the usual way.

We will refer to $x^{-1}\colon R_x \to U_x$ as a **chart at p** whenever $p \in U_x$.

Charts are just as important as coordinate maps.[1]

A chart x^{-1} inverse to coordinate map x can be used to "draw" curves through p on $\mathcal{M}$. First note that for each $v \in \mathbb{R}^n$ there is some largest $\varepsilon > 0$ so that $p_x + t\,v$ is in the range of x for every $t \in (-\varepsilon, \varepsilon)$. This allows us to define functions

$$\boldsymbol{G^v_{x,p}\colon (-\varepsilon,\varepsilon) \to \mathcal{M}} \quad \text{given by} \quad G^v_{x,p}(t) = x^{-1}(p_x + t\,v).$$



Of particular importance are the **coordinate gridcurves through p** on $\mathcal{M}$, the curves $G^{e_i}_{x,p}$ for $i = 1, \dots, n$. These curves trace out "bent" (by x) representations of the coordinate axis grid in $\mathbb{R}^n$ near p_x, which is taken to p by x^{-1}. Gridcurves are used as both tools for calculation and "visualization" aids. We will see them often.

The first three assumptions in the definition of manifold imply many other useful topological properties.

On a manifold, **sequences suffice to define continuity**: a function f defined on a manifold is continuous provided $f(t_n)$ is convergent to $f(p)$ in the range whenever t_n is a sequence that is convergent to p in the manifold. One can also show[2] that manifolds are **paracompact**, **locally compact**, **σ-compact**, **normal**, **locally path-connected**, of **second category** and **locally simply connected**. In a manifold, a **sequentially compact** subset is **compact**. **Path components** and **connected components** coincide in a manifold. Any two points in a path component can be connected by an arc: a one-to-one path. In addition, manifolds

[1]Note that not every source uses the words "chart" and "coordinate maps" as we do; these words may represent either x or x^{-1}.

[2]These various topological properties have consequences for us, and imply properties of manifolds which will be clearly stated at natural places in the text. We will focus on these important consequent properties, leaving most proofs that manifolds possess them for a class in point-set topology.

are **metrizable** and have a **partition of unity subordinate to any open cover.** Manifolds admit a **universal covering space.**

There are non-Hausdorff locally Euclidean spaces, so the Hausdorff condition is a legitimate additional requirement if we want metrizability.[3] An uncountable set with discrete topology is locally Euclidean and even metrizable (single points in such a space are mapped to the member of $\mathbb{R}^0 = \{0\}$) so second countability is also independent of the "locally Euclidean" condition.

There is a (nontrivial) result called the "**Invariance of Domain Theorem**" which we will use and assume.

Invariance of Domain: A continuous one-to-one function from an open subset of $\mathbb{R}^n$ into $\mathbb{R}^n$ has open range. This implies that such a function has continuous inverse on this range.

This has important implications for us.

First, because an open subset of $\mathbb{R}^m$ can be regarded as a *non-open* subset of $\mathbb{R}^n$ if $n > m$ there cannot be two coordinate maps at p onto open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, unless $m = n$.

To reiterate, if two coordinate maps x and y at p have ranges open in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively then $n = m$.

Suppose p and q are in the same connected component of $\mathcal{M}$. Suppose given coordinate maps x at p with range an open set in $\mathbb{R}^n$ and coordinate map y at q with range an open set in $\mathbb{R}^m$. There is a path in $\mathcal{M}$ connecting p to q and the image of this path is compact so there are a finite number of charts whose ranges cover this path. The dimension of the domain spaces must agree on the overlap of these domains all along the path, so again $n = m$.

This unique number n is called the **dimension** of the connected component of $\mathcal{M}$ to which p and q belong.

We add a fourth and final regularity property to our list of defining properties for topological manifolds: we insist that the dimension of each connected component should be the same.

A topological manifold $\mathcal{M}$ is a Hausdorff, second countable and locally Euclidean topological space whose dimension, already uniquely defined by these three properties for each connected component, does not vary between components.

The notation $\boldsymbol{dim(\mathcal{M}) = n}$ will be used to indicate that manifold $\mathcal{M}$ has dimension n, and $\mathcal{M}$ is called an **n-manifold** when it has dimension n.

A compact topological manifold is called a **closed manifold.**

Every point in a topological manifold has coordinate maps around that point onto the cube $(-1, 1)^n$, onto an open ball $B_r(0)$ around the origin of radius r

[3]Let X be the set $\mathbb{R} \cup \{*\}$ where the point $*$ is not a real number. The open sets in X are the usual open sets in $\mathbb{R}$ together with any set that can be obtained by taking one of these open sets which contains 0 and either adding $*$ to it or replacing 0 by $*$. So every point in this set has a neighborhood which is homeomorphic to $\mathbb{R}$. Still, this topology is T_1 but not T_2.

centered at 0 and a coordinate map onto all of $\mathbb{R}^n$. In each case a coordinate map may be found which sends any particular point to the origin in $\mathbb{R}^n$.

2. A Few Facts From Calculus

We will need some results from Calculus and discuss them here.

First is the **Inverse Function Theorem**.

2.1. ***Theorem.*** *Suppose S is an open subset of $\mathbb{R}^n$ and $f\colon S \to \mathbb{R}^n$ is continuously differentiable and the derivative matrix $f'(p)$ is nonsingular for some $p \in S$.*

Then there is an open subset A of S containing p so that:
(i) $f(A)$ is an open subset of $\mathbb{R}^n$ and
(ii) f is one-to-one on A and
(iii) f' is nonsingular on all of A and
(iv) $f^{-1}\colon f(A) \to A$ is differentiable with nonsingular derivative on $f(A)$ and
(v) $(f^{-1})'(y) = (f'(x))^{-1}$ whenever $x \in A$ and $y = f(x) \in f(A)$.

You will note that this theorem implies that the range $f(S)$ is a neighborhood in $\mathbb{R}^n$ of every point $f(x)$ for which $f'(x)$ is nonsingular.

The proof can be found in many Multi-Variable Calculus texts, including a very nice one in [5] Michael Spivak's "Calculus on Manifolds."

The proof of the **Implicit Function Theorem**, stated next, can be found in that source.

2.2. ***Theorem.*** *Suppose $f\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable in an open set containing the point $(t, s) \in \mathbb{R}^m \times \mathbb{R}^n$.*

Suppose also that the $m \times m$ matrix formed from the first m partial derivatives of the m coordinate functions $f^1, \ldots, f^m$ of f

$$M(y,x) = \big(D_j f^i(y,x) \big) \quad i,j = 1, \ldots, m$$

is nonsingular at, and hence in a neighborhood of, (t, s) and that $f(t, s) = 0$.

Then there is an open neighborhood A in $\mathbb{R}^n$ of s and an open neighborhood B in $\mathbb{R}^m$ of t so that for each $x \in A$ there is a unique $g(x) \in B$ for which $f(g(x), x) = 0$.

The function $g\colon A \to B$ is continuously differentiable.

Finally, we have facts about Taylor polynomials and their remainders.

Suppose $h\colon (a, b) \to \mathbb{R}$ is $n + 1$ times continuously differentiable and s and t are in (a, b). When $k \le n$ the degree-k **Taylor polynomial** for h centered at t is

$$P_k(s) = h(t) + h'(t)(s-t) + \cdots + \frac{h^{(k)}(t)}{k!}(s-t)^k.$$

This polynomial has derivatives that match h up to order k at t and is interpreted as a polynomial approximation to h which may be a good approximation for s near t. The **error or "remainder"** term $R_k(s) = h(s) - P_k(s)$ is given by

$$R_k(s) = \int_t^s \frac{h^{(k+1)}(u)}{k!}(s-u)^k du$$

a fact that is obviously true when $k = 0$. Assuming this formula is $R_k(s)$ for some $k < n$ an integration by parts shows that the formula holds for $k+1$ as well:

$$R_{k+1}(s) = \int_t^s \frac{h^{(k+2)}(u)}{(k+1)!}(s-u)^{k+1} du.$$

We conclude that the formula is valid up to $k = n$, and for all k if h is infinitely differentiable.

It is entirely possible that this remainder will remain large[4] no matter how big k is. That would make the Taylor polynomial a poor choice if your goal is to find a good approximation to h.

There is a higher-dimensional version of this theorem. Suppose $g\colon S \to \mathbb{R}$ has continuous partial derivatives of all orders up to $n+1$ on open set $S \subset \mathbb{R}^m$, and that p and q are in S and so is the entire line segment connecting p and q.

This segment can be parametrized by $l(s) = p + s(q-p)$ for $0 \le s \le 1$. Define $h\colon [0,1] \to \mathbb{R}$ by $h(s) = g(l(s))$. Note $h(1) = g(q)$ and $h(0) = g(p)$.

Also by the chain rule the first derivative of h is

$$\frac{d}{du}h(u) = \frac{d}{du}g(l(u)) = g'(l(u))l'(u) = \sum_{i=1}^m D_i g(l(u))(q^i - p^i)$$

and the second derivative, which will interest us later, is

$$\frac{d^2}{du^2}h(u) = \frac{d}{du}\sum_{i=1}^m D_i g(l(u))(q^i - p^i) = \sum_{i,j=1}^m D_{j,i} g(l(u))(q^i - p^i)(q^j - p^j).$$

Getting back to Taylor polynomials, we have

$$P_k(1) = g(p) + h'(0) + \cdots + \frac{h^{(k)}(0)}{k!}$$

and the error term is given by

$$R_k(1) = g(q) - P_k(1) = \int_0^1 \frac{h^{k+1}(u)}{k!}(1-u)^k du.$$

In the case of $k = 1$ we have

$$\begin{aligned} g(q) &= g(p) + h'(0) \ + \int_0^1 h''(u)(1-u)\,du \\ &= g(p) + \sum_{i=1}^m D_i g(p)(q^i - p^i) \\ &\qquad + \sum_{i,j=1}^m (q^i - p^i)(q^j - p^j) \int_0^1 D_{j,i} g(l(u))(1-u)\,du. \end{aligned}$$

And the last function, represented as a sum involving integrals, is continuously differentiable at q two fewer times than the differentiability assumption on g itself.

Note also that when q is near p the values of $D_{j,i}g(l(u))$ are all very nearly $D_{j,i}g(p)$ and so each integral is approximately $\frac{D_{j,i}g(p)}{2}$.

[4]Consider the function defined to be 0 at $t = 0$ and e^{-1/t^2} elsewhere. Every Taylor polynomial centered at $t = 0$ is the zero polynomial, so the remainder is always the function itself.

3. Differential Equations

Physicists set up theories about the world on manifolds, frequently embodied in differential equations that must be satisfied by parametrized paths in a manifold if they are to represent physical quantities that fall under the purview of the theory.

So for this user group, and for other reasons that will unfold later, we need to discuss differential equations; in particular we will consider here first order systems of linear differential equations.

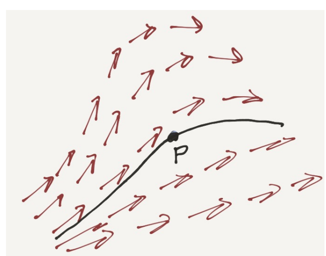
We suppose $\mathcal{U}$ is an open subset of $\mathbb{R}^n$ and (a, b) is an open interval which we may think of as a time parameter. If f is any function defined on (a, b) with values in $\mathbb{R}^k$ for some positive k we use $\dot{f}$ to denote the derivative of f, when it exists, with respect to the time parameter.

We suppose $g\colon (a, b) \times \mathcal{U} \to \mathbb{R}^n$ is any function.

A **differential equation** (DE for short) is an equation of the form

$$\dot{f}(t) = g(t, f(t)) \quad \text{for } t \in (a, b)$$

and a **solution to the differential equation** is a function $f\colon (r, s) \to \mathcal{U}$ that satisfies the differential equation for all t in some nonempty interval $(r, s) \subset (a, b)$.



An **initial value problem** (abbreviated IVP) is a differential equation as above with the additional requirement that the solution f must satisfy $f(t_0) = p$ for some chosen $t_0 \in (r, s)$ and specified $p \in \mathcal{U}$.

If the function g does not depend on its first domain variable (i.e. g is independent of time) the DE or IVP is called **autonomous** and if that is not being assumed the DE or IVP is called **nonautonomous**.

A nonautonomous equation can be converted to a related autonomous equation by the following subterfuge.

For nonautonomous IVP as above involving g define $h\colon (a, b) \times \mathcal{U} \to \mathbb{R}^{n+1}$ by $h(s, u) = (\,1, g(s, u)\,)$.

Consider a solution $w(t) = (t, f(t))$ to the autonomous IVP $\dot{w}(t) = h(w(t))$ with initial condition $w(t_0) = (t_0, p)$. Then f is a solution to the original IVP.

In all the cases we care about the function g will be continuous, which implies that any solution f will be continuously differentiable. In fact, if g is k times continuously differentiable then f will be at least $k + 1$ times continuously differentiable.

If g is merely continuous, the IVP can be reposed in operator form. Specifically, there is an interval (r, s) containing t_0 for which

$$f(t) = p + \int_{t_0}^{t} g(s, f(s))\, ds \quad \text{for } t \in (r, s).$$

The right hand side can be conceived of as an operator whose domain is a certain family of continuous functions

$$H(f)(t) = p + \int_{t_0}^{t} g(s, f(s))\, ds \quad \text{for } t \in (r, s)$$

and a solution to the IVP is a fixed point of this operator. The fixed point (and in fact any function produced by application of H to a continuous function f on an appropriate time interval) can be seen to be continuously differentiable.

In order to successfully deal with the domain issues for H and recast the problem in this form, and incidentally guarantee uniqueness of solutions, we need more conditions on g.

The function g is said to satisfy a **Lipschitz condition with Lipschitz constant L on $\mathfrak{U}$** provided $\| g(t, u) - g(t, q) \| \leq L \| u - q \|$ for every $u, q \in \mathfrak{U}$ and each t. *It is important to note that the number L does not depend on t.*

In case the closure of $(a, b) \times \mathfrak{U}$ is compact and g is defined and continuously differentiable on an open set containing this closure, the function g confined to $(a, b) \times \mathfrak{U}$ will satisfy a Lipschitz condition for *some* constant L.

3.1. ***Theorem. The Picard-Lindelöf Theorem*** *Suppose $\mathfrak{U}$ contains the closed ball B around p of radius K and (a, b) contains $[t_0 - T, t_0 + T]$ for positive T and K. Suppose further that g is continuous and the maximum value of $\|g\|$ on $[t_0 - T, t_0 + T] \times B$ is M and g satisfies a Lipschitz condition with Lipschitz constant L. Let R be the lesser of T or K/M.*

Then there is a unique solution to the IVP

$$\dot{f}(t) = g(t, f(t)) \quad \textit{for } t \in [t_0 - R, t_0 + R] \textit{ and } f(t_0) = p.$$

The proof of this theorem can be found in any differential equation text, such as the one by [6] Birkhoff and Rota, "Ordinary Differential Equations 4th Edition," and we outline a proof below.

The uniqueness condition implies that any other solution must agree with this one at all points in its domain that are also in $[t_0 - R, t_0 + R]$.

The statement of the theorem allows us to specify a domain for the operator form of the IVP. The solution whose existence is proposed in Theorem 3.1 will be the unique fixed point of this operator.

3.2. ***Corollary.*** *Under the same conditions as the preceding theorem, let C denote the set of all continuous functions of the form*

$$u \colon [t_0 - R, t_0 + R] \to B \quad \textit{with } u(t_0) = p.$$

Then $H(u) \in C$ for all $u \in C$.

Proof. $H(u)$ is continuous (in fact differentiable) for all t in $[t_0 - R, t_0 + R]$ since g is continuous and $(t, u(t))$ is always in the domain of g. We intend to show that $H(u) \in C$ whenever $u \in C$. Specifically, we need to show that $w(t) = H(u)(t)$ cannot escape B for $t \in [t_0 - R, t_0 + R]$.

$$\|w(t) - p\| = \left\| \int_{t_0}^{t} g(s, u(s))\, ds \right\| \leq \left| \int_{t_0}^{t} \|g(s, u(s))\|\, ds \right| \leq |t - t_0|\, M.$$

Since $|t - t_0| \leq R$ and R cannot exceed K/M we have $w(t) \in B$ for all t. □

You will note that the actual value of the Lipschitz constant is not mentioned in the theorem: only its *existence* is needed to guarantee a unique continuously differentiable solution in the specified regions. One of the proofs of Theorem 3.1, uses the **Picard iteration** scheme to produce successive approximations which converge uniformly to this unique solution; the Lipschitz constant is involved in estimating the *rate* of convergence of these approximations. We discuss this approach now.

Suppose g has Lipschitz constant L.

Define the supremum norm $\|\cdot\|_\infty$ on the Banach space of continuous $\mathbb{R}^n$-valued functions with domain $[t_0 - R, t_0 + R]$ by

$$\|u\|_\infty = \sup\{\, \|u(t)\| \mid t \in [t_0 - R, t_0 + R]\,\}.$$

The domain C of operator H given above is a complete metric space with the metric induced by this norm. We have, for each t,

$$\begin{aligned}\|H(u)(t) - H(w)(t)\| &= \left\| \int_{t_0}^{t} g(s, u(s)) - g(s, w(s))\, ds \right\| \\ &\leq \left| \int_{t_0}^{t} \|g(s, u(s)) - g(s, w(s))\|\, ds \right| \leq \left| \int_{t_0}^{t} L\, \|u(s) - w(s)\|\, ds \right| \\ &\leq L\, |t - t_0|\, \|u - w\|_\infty \leq L\, R\, \|u - w\|_\infty.\end{aligned}$$

So H is not necessarily a contraction on C, but it is Lipschitz with Lipschitz constant LR. The **Banach Fixed Point Theorem** would directly imply the existence of a unique fixed point for H, a solution to the IVP, only when H is a contraction: that is, when $LR < 1$.

There are two ways to proceed. The first is slightly less satisfactory than the second.

We could require that, in addition to our restriction that R be the lesser of T or K/M, that R be a positive number less than L. When C is defined using this (possibly) smaller time interval H will be a contraction using the metric from the supremum norm. The disadvantage is that our unique solution will be defined on a smaller symmetric interval centered at t_0, but for many purposes that will suffice.

A more satisfactory approach is to leave R alone but modify the metric to an equivalent metric for which H is actually a contraction on C. We define a new weighted norm $\|\cdot\|_*$ that suppresses magnitudes for times away from t_0 with

$$\|u\|_* = \sup\{\, e^{-2L|t - t_0|} \|u(t)\| \mid t \in [t_0 - R, t_0 + R]\,\}.$$

The weighting factor cannot exceed 1 and is always at least e^{-2LR} so we have

$$e^{-2LR} \|\cdot\|_\infty \leq \|\cdot\|_* \leq \|\cdot\|_\infty \leq e^{2LR} \|\cdot\|_*$$

so the same sets of functions are closed, and the same sequences of functions are Cauchy, when defined using the metric for either norm.

The values used to calculate $\|H(u) - H(w)\|_*$ are the numbers

$$
\begin{aligned}
e^{-2L|t-t_0|}\|H(u)(t) - H(w)(t)\| = \ & e^{-2L|t-t_0|} \left\| \int_{t_0}^{t} g(s,u(s)) - g(s,w(s))\, ds \right\| \\
&\leq e^{-2L|t-t_0|} \left| \int_{t_0}^{t} \|g(s,u(s)) - g(s,w(s))\|\, ds \right| \\
&\leq e^{-2L|t-t_0|} L \left| \int_{t_0}^{t} \|u(s) - w(s)\|\, ds \right| \\
&= e^{-2L|t-t_0|} L \left| \int_{t_0}^{t} \|u(s) - w(s)\| e^{-2L|s-t_0|} e^{2L|s-t_0|}\, ds \right| \\
&\leq e^{-2L|t-t_0|} L \left| \int_{t_0}^{t} \|u - w\|_*\, e^{2L|s-t_0|}\, ds \right| \\
&= e^{-2L|t-t_0|} L \|u-w\|_* \left| \int_{t_0}^{t} e^{2L|s-t_0|}\, ds \right| \\
&= e^{-2L|t-t_0|} L \|u-w\|_* \left| \frac{e^{2L|t-t_0|}}{2L} - \frac{1}{2L} \right| = \|u-w\|_* \left| \frac{1}{2} - \frac{e^{-2L|t-t_0|}}{2} \right| \\
&\leq \frac{\|u-w\|_*}{2}.
\end{aligned}
$$

So the supremum of all the original numbers, which is $\|H(u) - H(w)\|_*$, cannot exceed $\frac{\|u-w\|_*}{2}$. So with the metric from this norm H is a contraction with contraction constant $\varepsilon = 1/2$.

Let $u_0(t) = p$ for all $t \in [t_0 - R, t_0 + R]$ and, having found function u_j for $0 \leq j < k$ define, again for $t \in [t_0 - R, t_0 + R]$, the function u_k by

$$u_k(t) = H(u_{k-1}) = p + \int_{t_0}^{t} g(s, u_{k-1}(s))\, ds.$$

Applying the inequality from above we have, for $m > n > 1$,

$$\|u_m - u_n\|_* \leq \varepsilon \|u_{m-1} - u_{n-1}\|_* \leq \varepsilon^n \|u_{m-n} - u_0\|_* \leq \varepsilon^n 2K$$

and we conclude from this that the sequence of iterates is Cauchy and therefore converges uniformly to a function f in uniformly closed C.

Continuity of H implies $H(f) = \lim_{k\to\infty} H(u_k) = \lim_{k\to\infty} u_{k+1} = f$ so f is the fixed point we were after.

We have two useful inequalities governing the rate of convergence.

$$
\begin{aligned}
\|u_n - f\|_\infty \leq e^{2LR}\|u_n - f\|_* &\leq e^{2LR} \frac{\varepsilon^n}{1-\varepsilon} \|u_1 - u_0\|_* \\
&\leq e^{2LR} \frac{\varepsilon^n}{1-\varepsilon} \|u_1 - u_0\|_\infty
\end{aligned}
$$

and

$$\|u_n - f\|_\infty \leq e^{2LR}\|u_n - f\|_* \leq e^{2LR}\frac{\varepsilon}{1-\varepsilon}\|u_n - u_{n-1}\|_*$$
$$\leq e^{2LR}\frac{\varepsilon}{1-\varepsilon}\|u_n - u_{n-1}\|_\infty.$$

The first is an a priori estimate and can be calculated after the first Picard iterate is formed. If you have an idea of how close you must be to the desired solution this will tell you a "worst case" of how big n must be to accomplish that. This is usually a gross overestimate.

The second inequality[5] is more useful for purposes of estimation, and is called an a posteriori estimate by those in the approximation business. It is calculated as you proceed and tells you when you can stop creating new u_k.

We consider now the situation of an autonomous differential equation.

Initial value problems may be rephrased as the search for solution curves c for which $\dot{c} = g \circ c$ and $c(t_0) = p$.

In this autonomous case the solution to the related IVP $\dot{b} = g \circ b$ and $b(t_1) = p$ is given by $b(t) = c(t_0 + t - t_1)$ for $t \in \left[t_1 - \frac{K}{M}, t_1 + \frac{K}{M}\right]$.

So because of the uniqueness condition in the Picard-Lindelöf Theorem, changing the initial time at which the solution curve passes through a point does not change the solution curve beyond a time-shift. The corresponding feature *does not* hold when the IVP is nonautonomous. The restriction on the time interval[6] from that theorem, $t \in [t_0 - K/M,\ t_0 + K/M]$, is to prevent the solution curve from leaving the ball B of radius K, and thereby (possibly) leaving the domain $\mathcal{U}$ of Lipschitz g.

We assume that g is Lipschitz with Lipschitz constant L and maximum magnitude M on the domain of g as in Theorem 3.1 and Corollary 3.2 in this autonomous case.

Let $\boldsymbol{B_q(r)}$ denote the closed ball in $\mathbb{R}^n$ of radius r centered at q and $\boldsymbol{B_q^o(r)}$ the interior of that ball.

So the ball B specified in those results is $B_p(K) \subset \mathcal{U}$ and the time interval upon which the solution is defined and for which the curve, starting at p when $t = 0$, will not leave $B_p(K)$ is $\left[-\frac{K}{M},\ \frac{K}{M}\right]$. That same curve will never leave ball $B_p(K/2)$ on time interval $\left[-\frac{K}{2M},\ \frac{K}{2M}\right]$.

And from each $q \in B_p(K/2)$ we have $B_q(K/2) \subset B_p(K)$. So the theorem and corollary apply for every $q \in B_p(K/2)$ and yield solution curves

$$c_q\colon \left[-\frac{K}{2M},\ \frac{K}{2M}\right] \to B_p(K) \quad \text{and} \quad \dot{c_q} = g \circ c_q \quad \text{and} \quad c_q(0) = q.$$

Sometimes it is the curves themselves that are of interest, but sometimes a slight change in notation is useful to emphasize a family of functions on the ball with a

[5]Prove this one first, using $\|u_n - f\|_* \leq \varepsilon\|u_{n-1} - f\|_* = \varepsilon\|u_{n-1} - u_n + u_n - f\|_*$ and the triangle inequality.

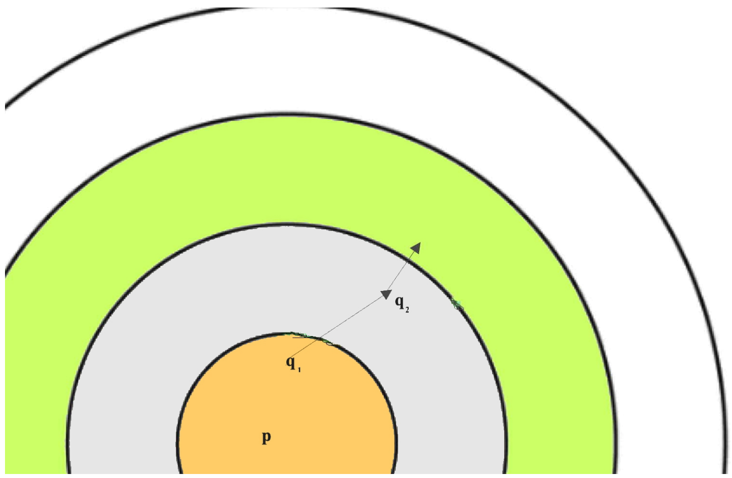
[6]We do want the time interval to be bounded in order to save us from technical difficulties in the convergence result. Conclusions for an unbounded time interval can be drawn by taking limits of intervals of increasing but finite size if the domain of g is unbounded.

time parameter. In that case we use

$$S\colon \left[-\frac{K}{2M},\ \frac{K}{2M}\right] \times B_p(K/2) \to B_p(K) \text{ given by } S_t(q) = c_q(t).$$

S is called a **local flow for the DE**. For each fixed time t the function S_t sends points in $B_p(K/2)$ to points in $B_p(K)$.

Starting from any $q_1 \in B_p(K/4)$ and traveling on a solution curve (in a positive or negative direction) for time less than $\frac{K}{4M}$ you can't leave $B_p(K/2)$. If you end up at point $q_2 \in B_p(K/2)$ after this journey from q_1 you *still* can't leave $B_p(K)$ during an *additional* time $\frac{K}{4M}$—in fact you will have time $\frac{K}{4M}$ to spare before you could get near the boundary of $B_p(K)$.



With this in mind (so the following line is defined) and invoking the uniqueness of solutions to these IVPs we see that for any $q \in B_p(K/4)$ and any $t, s \in [-\frac{K}{4M}, \frac{K}{4M}]$ we have

$$S_{s+t}(q) = S_t(c_q(s)) = S_t(S_s(q)).$$

Rephrasing the differentiation properties in this new notation, for q in $B_p^o(K/2)$

$$S_0(q) = q \quad \text{and} \quad \lim_{\varepsilon \to 0} \frac{S_\varepsilon(q) - q}{\varepsilon} = g(q).$$

Also, for all $q \in B_p^o(K/4)$ and $s \in [-R/4, R/4]$

$$\lim_{\varepsilon \to 0} \frac{S_{\varepsilon+s}(q) - S_s(q)}{\varepsilon} = g(S_s(q)).$$

So for each fixed $t \in (-R/4, R/4)$ we have a function $S_t\colon B_p^o(K/4) \to B_p^o(K/2)$ and this function has inverse function which can be calculated using S_{-t}. Therefore it must be one-to-one on $B_p^o(K/4)$.

Suppose q_1 and q_2 are in the interior of $B_p(K/4)$ and $t \in (0, R/4)$. Using the integral operator form of the IVP we have

$$S_t(q_1) - S_t(q_2) = q_1 - q_2 + \int_0^t g(S_s(q_1)) - g(S_s(q_2))\, ds.$$

The continuous function $f(s) = \| S_s(q_1) - S_s(q_2) \|$ for $s \in [0,t]$ satisfies the condition

$$f(r) \leq f(0) + \int_0^r L\, f(s)\, ds$$

where L is the Lipschitz constant for g so by the integral form of **Grönwall's Inequality** we have

$$f(t) \leq f(0)\, e^{\int_0^t L\, ds} = f(0)\, e^{tL}.$$

That means the magnitude of $S_t(q_1) - S_t(q_2)$ can never exceed $\| q_1 - q_2 \| \, e^{tL}$ which puts an upper bound on how fast neighboring solution curves can separate from each other.

And in particular we see that each S_t is continuous.

Invariance of domain tells us the image set $S_t(B_p^o(K/4))$ must be open, and S_t itself is a homeomorphism between $B_p^o(K/4)$ and this image with inverse S_{-t}.

Proving differentiability of the local flow[7] is quite a bit more involved than continuity and we leave that discussion for elsewhere, for instance [4] Lang's "*Introduction to Differentiable Manifolds 2nd Ed.*" page 75.

There we see that if g is C^k so is any local flow determined by g.

4. Differentiable Manifolds

A $\mathbf{C^k}$ **atlas** on a topological manifold $\mathcal{M}$ is a set of coordinate maps $\mathcal{A}$ of $\mathcal{M}$ whose domains cover $\mathcal{M}$ and with the property that whenever $x\colon U_x \to R_x$ and $y\colon U_y \to R_y$ are coordinate maps from $\mathcal{A}$

$$y^i \circ x^{-1}\colon x(\, U_x \cap U_y\,) \subset \mathbb{R}^n \to \mathbb{R}$$

has continuous mixed partial derivatives up to order k for each coordinate y^i.

We assume below that x and y are members of a C^k atlas where $k \geq 1$.

If f is any real valued function defined on a neighborhood of $p \in \mathcal{M}$ (such as the y^i above) we define

$$\boldsymbol{\frac{\partial f}{\partial x^j}(p)} \quad \text{to be} \quad D_j(f \circ x^{-1})(p_x) = \left[\, D_j(f \circ x^{-1}) \,\right] \circ x(p)$$

where D_j denotes partial differentiation with respect to the jth coordinate in $\mathbb{R}^n$.

Numerically, this is the rate of change of f as you move on $\mathcal{M}$ through p along the jth coordinate gridcurve for x at unit pace: that is, with respect to the natural parameterization $G_{x,p}^{e_j}(t) = x^{-1}(p_x + t\, e_j)$.

Higher partial derivatives will be needed later, and are defined as you would expect, according to the pattern

$$\begin{aligned}\boldsymbol{\frac{\partial^2 f}{\partial x^k \partial x^j}(p)} &= \frac{\partial \frac{\partial f}{\partial x^j}}{\partial x^k}(p) = D_k \left(\left[\, D_j(f \circ x^{-1}) \,\right] \circ x \circ x^{-1} \right)(p_x)\\ &= (D_k D_j f \circ x^{-1})(p_x).\end{aligned}$$

[7]We have, of course, differentiability of the solution curves that, collectively, form the flow.

The **Jacobian matrices** $\boldsymbol{\frac{dx}{dy}(p)}$ and $\boldsymbol{\frac{dy}{dx}(p)}$ corresponding to the two charts at $p \in U_x \cap U_y$ are given by

$$\boldsymbol{\frac{dy}{dx}(p) = (y \circ x^{-1})'(p_x)} = \left(\frac{\partial y^i}{\partial x^j}(p) \right)$$

and

$$\boldsymbol{\frac{dx}{dy}(p) = (x \circ y^{-1})'(p_y)} = \left(\frac{\partial x^i}{\partial y^j}(p) \right).$$

The Jacobian matrix $(y \circ x^{-1})'(p_x)$, acting by left matrix multiplication on $\mathbb{R}^n$, is the linear map that best approximates $y \circ x^{-1}$ near p_x in the sense that this is the unique matrix J for which

$$\lim_{q \to p} \frac{\left\| y \circ x^{-1}(q_x) - y \circ x^{-1}(p_x) - J\,(q_x - p_x) \right\|}{\|\, q_x - p_x \,\|} = 0.$$

The Jacobian matrix $(x \circ y^{-1})'(p_y)$ has full rank at every $p_y \in y(\,U_x \cap U_y\,)$ and $(\,(x \circ y^{-1})'(p_y)\,)^{-1} = (y \circ x^{-1})'(p_x)$ for every $p \in U_x \cap U_y \subset \mathcal{M}$.

On a manifold itself and neighboring points p and q it makes no sense to talk about, in general, a displacement "vector" $q - p$ from p to q since no vector operations are defined. What *does* make sense are the **coordinate displacements** $h_x = q_x - p_x$ and $h_y = q_y - p_y$. The equation above then reads

$$\lim_{q \to p} \frac{\| h_y - J\, h_x \|}{\|\, h_x \,\|} = 0$$

and this tells us that on the overlap of two coordinate patches we can use coordinate displacements to discuss actual (small) displacements on the manifold itself, translating this information between coordinate maps using the Jacobian. The error you make by doing this for a given coordinate displacement is small *even in comparison to the coordinate displacement* when q is near p.

A $\mathbf{C^k}$ **differentiable structure** is a maximal $\mathbf{C^k}$ **atlas**: that is, an atlas to which no additional charts can be added while retaining the differentiability condition. A differentiable structure containing an atlas is said to be **compatible** with the atlas, and there is one and only one $\mathbf{C^k}$ differentiable structure compatible with any particular $\mathbf{C^k}$ atlas.

We note that if $x \colon U_x \to R_x \subset \mathbb{R}^n$ is a coordinate map in a differentiable structure so is the restriction $x|_V$ of x to any open subset V of U_x, and this is an observation we will use from time to time.

A C^∞ manifold is called, simply, a **differentiable or smooth manifold** or, if compact, a **closed differentiable manifold.**

From now on all our manifolds will be differentiable and endowed with some specific $\boldsymbol{C^\infty}$ differentiable structure.

That is equivalent to requiring that the real valued function $y^i \circ x^{-1}$ has continuous mixed partials of all orders for every $i = 1, \ldots, n$ at every point in $x(\,U_x \cap U_y\,)$ and every pair of coordinate maps x and y in the differentiable structure or, equivalently, in a generating atlas.

$\mathbb{R}^{\boldsymbol{n}}$ **will be presumed to have the differentiable structure consisting of all coordinate maps compatible in this way with the identity map:** that is, all coordinate maps x for which x and x^{-1} are differentiable in the usual sense to all orders on their domains.

Suppose given a function $f\colon \mathcal{O} \to \mathcal{N}$ where $\mathcal{O}$ is an open subset of differentiable n-manifold $\mathcal{M}$ and that $\mathcal{N}$ is a differentiable k-manifold.

f is called **differentiable** if $y\circ f\circ x^{-1}$ is differentiable for every chart $x^{-1}\colon R_x \to U_x \subset \mathcal{M}$ and any coordinate map $y\colon U_y \to R_y \subset \mathbb{R}^k$ for $\mathcal{N}$ wherever this composition is defined, specifically for every $p_x \in x\left(U_x \cap f^{-1}(U_y)\right)$.

Differentiability of $y \circ f \circ x^{-1}$ implies continuity so $x\left(U_x \cap f^{-1}(U_y)\right)$ is an open subset of R_x.

To check differentiability of f it is not necessary to work with *every* chart in each differentiable structure. Looking at those charts in any pair of atlases that *generate* the relevant differentiable structures is sufficient.

In a generalization of the previous definition, the **derivative or Jacobian matrix** of f with respect to coordinate maps x and y is

$$\boldsymbol{\frac{d\,y\circ f}{dx}(p) = (y\circ f\circ x^{-1})'(p_x)} = \left(\frac{\partial(y^i\circ f)}{\partial x^j}(p)\right), \qquad p\in U_x\cap f^{-1}(U_y).$$

This is the unique matrix M, when these partial derivatives exist and are continuous, for which

$$\lim_{q\to p}\frac{\left\| y\circ f\circ x^{-1}(q_x) - y\circ f\circ x^{-1}(p_x) - M\,(q_x - p_x)\right\|}{\|q_x - p_x\|} = 0.$$

The **rank of f at p** $\in U_x \cap f^{-1}(U_y)$ is the rank of its Jacobian matrix. Rank does not depend on the particular charts used to define it.

A Jacobian matrix for two coordinate maps is, of course, the derivative matrix with respect to these coordinate maps of the identity function on the manifold.

A function $f\colon U \to \mathcal{N}$ defined on an open subset U of $\mathcal{M}$ is called $\mathbf{C^k}$ (also, **continuously differentiable** if $k = 1$, or **smooth** if $k = \infty$) if $y^i \circ f \circ x^{-1}$ is differentiable with continuous partials up to order k for all members y and x of the relevant atlases on $\mathcal{N}$ and $\mathcal{M}$ respectively and all coordinates y^i at each point for which the composite function is defined. These points are all in $(U_x \cap f^{-1}(U_y)$, which may of course be empty for any specific pair of x and y. However every point in U is in such a set for *some* x and y.

This definition coincides with the usual one when $\mathcal{N}$ is an open subset of $\mathbb{R}^k$ and $\mathcal{M}$ is an open subset of $\mathbb{R}^n$.

The set of real valued functions which are C^k and defined on all of $\mathcal{M}$ is denoted $\boldsymbol{\mathcal{F}^k(\mathcal{M})}$. It is a real commutative algebra with unit, infinite dimensional except in trivial cases.

The set of real valued functions that are defined and C^k on some neighborhood of $q \in \mathcal{M}$ will be denoted $\boldsymbol{\mathcal{F}^k_q(\mathcal{M})}$.

The cases $k = 1$ and $k = \infty$ will be the only ones that concern us. When the manifold $\mathcal{M}$ is fixed and understood we will forgo $\mathcal{F}^k(\mathcal{M})$ and $\mathcal{F}^k_q(\mathcal{M})$ in favor of $\mathcal{F}^k$ and $\mathcal{F}^k_q$.

$\mathcal{F}_q^k$ also is a real commutative algebra with unit when we define fg and $f + rg$ in the obvious way on the intersection of the domains of f and g, for real r and $f, g \in \mathcal{F}_q^k$.

A function $f \colon \mathcal{M} \to \mathcal{N}$ is called a **diffeomorphism**, and the two C^∞ manifolds $\mathcal{M}$ and $\mathcal{N}$ are called **diffeomorphic**, if f is one-to-one and onto and y is in the differentiable structure for $\mathcal{N}$ if and only if $y \circ f$ is in the differentiable structure for $\mathcal{M}$.

Thus, a diffeomorphism f is C^∞, the dimensions of diffeomorphic manifolds are the same, and if $f \colon \mathcal{M} \to \mathcal{N}$ is a diffeomorphism, so is $f^{-1} \colon \mathcal{N} \to \mathcal{M}$.

For example, consider $\mathbb{R}$ with standard differentiable structure and the interval $(-1, 1)$ with *its* standard structure inherited from $\mathbb{R}$ and the map

$$f \colon (-1, 1) \to \mathbb{R} \quad \text{given by} \quad x \to \tan\left(\frac{\pi x}{2}\right).$$

If a function $g \colon U \subset \mathbb{R} \to \mathbb{R}$ is differentiable a certain number of times (this includes coordinate maps for $\mathbb{R}$) then so is $g \circ f$ by the chain rule. And if $h \colon U \subset (-1, 1) \to \mathbb{R}$ is differentiable a certain number of times, so is $h \circ f^{-1}$ for the same reason.

It follows that $\mathbb{R}$ and $(-1, 1)$ with these standard differentiable structures are diffeomorphic and $f \colon (-1, 1) \to \mathbb{R}$ is a diffeomorphism demonstrating this.

Now give $\mathcal{M} = \mathbb{R}$ the standard differentiable structure but give $\mathcal{N} = \mathbb{R}$ the differentiable structure compatible with the atlas containing the single function g given by $g(x) = x^3$. Note that the identity map $x \to x$ is *not* in the differentiable structure for $\mathcal{N}$ because g^{-1} is not differentiable at 0.

Coordinate maps whose domains do not contain 0 in $\mathcal{N}$ coincide with those in the standard differentiable structure, but if y is a coordinate map around 0 in $\mathcal{N}$ we must have $h = y \circ g^{-1}$ differentiable. That means y is of the form $h \circ g$ for differentiable h. And of course if h is any differentiable function so is $h \circ g$.

So $g \colon \mathcal{M} \to \mathcal{N}$ is a diffeomorphism.

Finally, if $\mathcal{M}$ and $\mathcal{N}$ are smooth manifolds of dimensions n and k respectively, we define the smooth manifold $\boldsymbol{\mathcal{M} \times \mathcal{N}}$, called a **product manifold**, to be the obvious set of ordered pairs with product topology together with the differentiable structure compatible with the atlas of functions $(x, y) \colon U_x \times U_y \to \mathbb{R}^n \times \mathbb{R}^k = \mathbb{R}^{n+k}$ given by $(x, y)(c, d) = (x(c), y(d))$ where x is in the differentiable structure for $\mathcal{M}$ and y is in the differentiable structure for $\mathcal{N}$.

With this definition, the **projection maps**

$$\boldsymbol{\pi_1} \colon \mathcal{M} \times \mathcal{N} \to \mathcal{M} \quad \text{and} \quad \boldsymbol{\pi_2} \colon \mathcal{M} \times \mathcal{N} \to \mathcal{N}$$

given by $\pi_1(c, d) = c$ and $\pi_2(c, d) = d$ are automatically smooth.

5. Partitions of Unity

If X is a topological space, the **support** of any function $f\colon X \to \mathbb{R}$, denoted **supp(f)**, is the closure of the set $f^{-1}(\mathbb{R} - \{0\})$.

In a topological space X a **cover** of a set $S \subset X$ is a set $\boldsymbol{\mathcal{A}}$ of sets whose union contains S. The cover $\boldsymbol{\mathcal{A}}$ is called **open** if every set in $\boldsymbol{\mathcal{A}}$ is open. The cover $\boldsymbol{\mathcal{B}}$ is said to be a **refinement** of the cover $\boldsymbol{\mathcal{A}}$ if every member of $\boldsymbol{\mathcal{B}}$ is contained in a member of $\boldsymbol{\mathcal{A}}$. A cover is called **locally finite** if only finitely many of the sets in the cover contain any particular point of S.

A **partition of unity** on a manifold $\mathcal{M}$ is a set $\mathcal{H}$ of continuous functions with range contained in $[0, 1]$ and for which $\{\, supp(f) \mid f \in \mathcal{H}\,\}$ is a locally finite cover of $\mathcal{M}$, and for which $\sum_{f\in\mathcal{H}} f(p) = 1$ for every $p \in \mathcal{M}$.

The partition of unity is said to be **subordinate** to a cover $\boldsymbol{\mathcal{A}}$ of $\mathcal{M}$ if the supports of the members of the partition of unity form a refinement of $\boldsymbol{\mathcal{A}}$.

The **Existence of Subordinate Partitions of Unity Theorem** states that there is a partition of unity subordinate to *any* given open cover of topological manifold $\mathcal{M}$. There is an open cover consisting of open sets with compact closure which refines any given open cover. So the partition of unity can be chosen to consist of functions with compact support subordinate to any given open cover.
If the manifold is differentiable, the member functions of this partition of unity can be chosen to be smooth.

Let $B^o_r(s) = \{\, m \in \mathbb{R}^n \mid \|m - s\| < r\,\}$ denote the open ball in $\mathbb{R}^n$ centered at s of radius r.

Suppose $\boldsymbol{\mathcal{K}}$ is any open cover of differentiable manifold $\mathcal{M}$. There is a countable set of charts $Q_i\colon B^o_3(0) \to U_i$, $i = 1, 2, \ldots,$ for which $\boldsymbol{\mathcal{U}}$, consisting of the sets $U_i = Q_i(B^o_3(0))$ forms a locally finite open cover of $\mathcal{M}$ which refines $\boldsymbol{\mathcal{K}}$.

Further, a differentiable atlas of this kind can be found so that $\boldsymbol{\mathcal{V}}$ and $\boldsymbol{\mathcal{W}}$, consisting of the sets $V_i = Q_i(B^o_2(0))$ and $W_i = Q_i(B^o_1(0))$ respectively, also form locally finite open covers of $\mathcal{M}$ so they too are refinements of $\boldsymbol{\mathcal{K}}$.

In the proof of this theorem, a countable set of smooth functions $g_i\colon M \to [0, 1]$ are constructed for which $W_i \subset g_i^{-1}(1)$ and $supp(g_i)$ is contained in the closure of V_i for each i.

Now the functions

$$f_i = \frac{g_i}{\sum_{k=1}^{\infty} g_k} \qquad i = 1, 2, \ldots$$

form the partition of unity subordinate to $\boldsymbol{\mathcal{K}}$, which we sought in the statement of the theorem.

There is a useful corollary, or special case of the theorem that deserves mention.

If $\mathcal{M}$ is a topological manifold and $\varnothing \neq C \subset V \subset M$ and V is open and C is compact then there is a function f defined on $\mathcal{M}$ with f identically 1 on C so that f has compact support contained in V and for which the set $\{f, 1-f\}$ is a partition of unity subordinate to the open cover $\{V,\ \mathcal{M}-C\}$ of $\mathcal{M}$. Further, if $\mathcal{M}$ is differentiable, f can be chosen to be smooth.

Suppose F is any real valued function defined on an open set $V \subset \mathcal{M}$ and C is a compact subset of V. Then there is a function G defined on all of $\mathcal{M}$ with $supp(G) \subset V$ that agrees with F on C and with $G \leq F$ on V. G can be chosen to match the differentiability of F.

Function G is defined to be Ff on V and 0 off V, where f is given in the boxed comment above. Thus, any function defined on a neighborhood of a point p in $\mathcal{M}$ is equal on a (possibly smaller) neighborhood of p to a function defined on all of $\mathcal{M}$, with matching differentiability properties.

So every member of $\boldsymbol{\mathcal{F}_p^k}$ agrees with a member of $\boldsymbol{\mathcal{F}^k}$ on a neighborhood of p.

6. Rank and Some Special Charts

In this section we mostly state results without proof typically found (following from the implicit function theorem) in Advanced Calculus. A good source for proofs is [5] Spivak's "Calculus on Manifolds" or [7] Loomis and Sternberg's very good "Advanced Calculus 2nd Edition", although any book at this level is likely to deal with these topics.

A set $S \subset \mathbb{R}^n$ is said to be **measure zero** if for each $\varepsilon > 0$ there is a countable set $C_n,\ n = 1, 2, \ldots,$ of cubes in $\mathbb{R}^n$ for which

$$S \subset \bigcup_{i=1}^{\infty} C_i \qquad \text{and} \qquad \sum_{i=1}^{\infty} Vol(C_i) < \varepsilon$$

where $Vol(C_i)$ denotes the ordinary volume of the cube C_i.

A subset S of a differentiable manifold $\mathcal{M}$ is said to be **of measure zero** if there is a countable sequence of coordinate maps $x_i \colon U_i \to R_i \subset \mathbb{R}^n$ for which $x_i(S \cap U_i)$ is of measure zero for each i and S is contained in the union of these U_i.

6.1. ***Theorem.*** *If $f \colon \mathbb{R}^n \to \mathbb{R}^n$ is C^1 and $S \subset \mathbb{R}^n$ has measure 0 then $f(S)$ has measure 0.*

This immediately implies the following:

If $f \colon \mathcal{M} \to \mathcal{N}$ is a C^1 map between two n-dimensional manifolds and $S \subset \mathcal{M}$ has measure 0 then $f(S)$ has measure 0.

Suppose $f \colon \mathcal{M} \to \mathcal{N}$ where $dim(\mathcal{M}) = n$ and $dim(\mathcal{N}) = k$.

A point $p \in \mathcal{M}$ is called a **critical point** of f if f has rank less than k, the dimension of its range $\mathcal{N}$. Since the rank cannot exceed n, if $n < k$ every point of $\mathcal{M}$ is automatically critical.

The **critical values** of f are defined to be the set of those $f(p) \in \mathcal{N}$ for some critical point p of f. **Regular values** of f are members of $\mathcal{N}$ which are *not* critical values, and that includes all points in $\mathcal{N} - f(\mathcal{M})$.

We have **Sard's Theorem:**

6.2. ***Theorem.*** *If $f\colon \mathcal{M} \to \mathcal{N}$ is a C^1 map between two n-dimensional manifolds the set of critical values of f has measure 0.*

6.3. ***Theorem.*** *Suppose that $f\colon \mathcal{M} \to \mathcal{N}$ is a C^1 map between n-manifold $\mathcal{M}$ and k-manifold $\mathcal{N}$. So the rank of f can exceed neither n nor k.*

(i) If f has rank $j < k$ at w then there are coordinate maps x around w and y around $f(w)$ and differentiable functions $\psi^{j+1}, \ldots, \psi^k$ such that

$$y \circ f \circ x^{-1} \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^j \\ \psi^{j+1}(a) \\ \vdots \\ \psi^k(a) \end{pmatrix} \quad \textit{for all} \quad a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \quad \textit{near} \quad x(w).$$

(ii) If f has rank $j < k$ in a neighborhood of w then there are coordinate maps x around w and y around $f(w)$ such that

$$y \circ f \circ x^{-1} \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^j \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \textit{for all} \quad a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \quad \textit{near} \quad x(w).$$

(iii) If f has rank k at w then there are coordinate maps x around w and y around $f(w)$ such that

$$y \circ f \circ x^{-1} \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^k \end{pmatrix} \quad \textit{for all} \quad a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \quad \textit{near} \quad x(w).$$

In case (ii) above, x and y exist so that f takes the first j coordinate gridcurves through w for x onto the first j coordinate gridcurves through $f(w)$ for y, and the parameters for these gridcurves coincide, and f is constant on the last $n-j$ of these gridcurves for x.

A function $f\colon \mathcal{M} \to \mathcal{N}$ between n-manifold $\mathcal{M}$ and k-manifold $\mathcal{N}$ is called a **topological immersion** if it is continuous and *locally* one-to-one: that is, if there is a neighborhood of every point in the domain upon which f is one-to-one. If a topological immersion is differentiable and of rank n (the dimension of the domain) everywhere (so we must have $k \geq n$) it is called, simply, an **immersion.**

A subset $\mathcal{K}$, with its own differentiable structure, of manifold $\mathcal{N}$ is called an **immersed submanifold** if the **inclusion map $\boldsymbol{i}\colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{N}}$** is an immersion. This does *not* imply that coordinate maps on $\mathcal{K}$ can, necessarily, be extended to coordinate maps on $\mathcal{N}$, and the topology on $\mathcal{K}$ might be finer than the relative topology it inherits as a subset of $\mathcal{N}$.

6.4. ***Theorem.*** *If $f\colon \mathcal{M} \to \mathcal{N}$ is smooth and $\mathcal{K}$ is an immersed submanifold of $\mathcal{N}$ and $f(\mathcal{M}) \subset \mathcal{K}$ then if f is continuous considered as a map into $\mathcal{K}$ it is also smooth considered as a map into $\mathcal{K}$.*

If $f\colon \mathcal{M} \to \mathcal{N}$ is an immersion and a homeomorphism onto its image (so, in particular, it is globally one-to-one) it is called an **embedding.**

An immersed submanifold $\mathcal{K}$ of $\mathcal{N}$ is called a **submanifold** if the inclusion map $i\colon \mathcal{K} \to \mathcal{N}$ is an embedding. In this case the topology on $\mathcal{K}$ *is* the relative topology.

It is obvious that an open subset of a smooth n-manifold $\mathcal{M}$ can be regarded as a submanifold of the same dimension, using differentiable structure obtained by restricting domains of coordinate maps on $\mathcal{M}$ to V. These are called **open submanifolds.**

Submanifolds of lower dimension are more interesting in some ways, but no submanifold is more useful than an open subset $\mathcal{O}$ of our paradigmatic manifold $\mathbb{R}^n$ with differentiable structure from the identity map. In that case the identity map on $\mathcal{O}$ is a diffeomorphism of $\mathcal{O}$ onto its image in $\mathbb{R}^n$.

If $f\colon \mathcal{M} \to \mathcal{N}$ is smooth and if, for each $p \in \mathcal{M}$, there are open submanifolds $A_p \subset \mathcal{M}$ and $B_{f(p)} \subset \mathcal{N}$ which contain p and $f(p)$, respectively, for which the restriction of f to A_p is a diffeomorphism onto $B_{f(p)}$ we call f a **local diffeomorphism**. These are an important type of immersion.

If submanifold $\mathcal{K}$ is closed in $\mathcal{M}$ it is called a **closed submanifold of $\mathcal{M}$**.[8] If $\mathcal{M}$ has multiple components, each component is a closed submanifold of $\mathcal{M}$.

6.5. ***Theorem.*** *Suppose n-manifold $\mathcal{M}$ is a submanifold of k-manifold $\mathcal{N}$ and $x\colon U_x \to R_x \subset \mathbb{R}^n$ is a coordinate map for $\mathcal{M}$ around p.*

Then there is a coordinate map $y\colon U_y \to R_y \subset \mathbb{R}^k$ for $\mathcal{N}$ around p for which

$$y(q) = \begin{pmatrix} y^1(q) \\ \vdots \\ y^k(q) \end{pmatrix} = \begin{pmatrix} x^1(q) \\ \vdots \\ x^n(q) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \textit{whenever} \quad q \in \mathcal{M} \cap U_y.$$

y can be chosen so that if $q \in U_y$ and $y^i(q) = 0$ for $i = k+1, \ldots, n$ then $q \in \mathcal{M}$.

Note we are not saying that the set $\mathcal{M} \cap U_y$ is all of U_x, but it is a relatively open neighborhood of p in $\mathcal{M}$ and a subset of U_x.

[8]Recall that a closed manifold is by definition compact so, oddly, a closed submanifold of $\mathcal{M}$ need not itself be a closed manifold. This is unfortunate but standard terminology.

Thus an atlas for $\mathcal{M}$ can be formed which produces the differentiable structure for $\mathcal{M}$ and which is composed entirely of the restrictions of coordinate maps such as x to $\mathcal{M} \cap U_y$, selecting one such restriction for every p and x pair, where $x\colon U_x \to R_x$ is in the differentiable structure for $\mathcal{M}$ and $p \in U_x$.

6.6. ***Theorem.*** *Suppose $f\colon \mathcal{M} \to \mathcal{N}$, for n-manifold $\mathcal{M}$ and k-manifold $\mathcal{N}$, is smooth and p is in the range of f.*

If f has rank m on an open set containing $f^{-1}(p)$ then $f^{-1}(p)$ is a closed submanifold of $\mathcal{M}$ of dimension $n-m$.

So if p is not a critical value of f then $f^{-1}(p)$ is a closed submanifold of dimension $n-k$.

6.7. ***Theorem.*** *Suppose $f\colon \mathcal{M} \to \mathbb{R}$ is smooth, $\mathcal{M}$ is a closed submanifold of $\mathcal{N}$ and U is an open subset of $\mathcal{N}$ containing $\mathcal{M}$.*

Then f can be extended to a function defined on all of $\mathcal{N}$ whose support is contained in U.

7. The Tangent Space

For the next two sections we focus on a single chart around p in differentiable n-manifold $\mathcal{M}$ with coordinates $x\colon U_x \to R_x \subset \mathbb{R}^n$.

An important case arises when $\mathcal{M}$ is an open subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. Polar coordinates in $\mathbb{R}^2$ and spherical and cylindrical coordinates in $\mathbb{R}^3$ provide examples here.

Another important case arises when $\mathcal{M}$ lies in $\mathbb{R}^{n+1}$ and is the graph of a differentiable function $h\colon \mathcal{O} \to \mathbb{R}$ where $\mathcal{O}$ is open in $\mathbb{R}^n$. In that event we could use coordinates $x\colon \mathcal{M} \to \mathcal{O}$ given by

$$x\begin{pmatrix} w^1 \\ \vdots \\ w^n \\ h(w) \end{pmatrix} = \begin{pmatrix} w^1 \\ \vdots \\ w^n \end{pmatrix}.$$

Keep these examples in mind.

Now suppose $h\colon (a,b) \to \mathcal{M}$ is continuously differentiable, where (a,b) is a nonempty interval.

h is called a **differentiable parameterization of a curve through p** provided $h(\alpha) = p$ for some $\alpha \in (a,b)$. We know there are differentiable parameterizations through p. The gridcurves $G^v_{x,p}$ are infinitely differentiable examples.

$$G^v_{x,p}(0) = x^{-1}(p_x) = p \quad \text{and} \quad (x \circ G^v_{x,p})'(0) = (p_x + t\,v)'(0) = v.$$

So any vector in $\mathbb{R}^n$ can be obtained as $(x\circ h)'(0)$ for *some* infinitely differentiable parameterization through p, and $G^v_{x,p}$ is the simplest one.

For each $v \in \mathbb{R}^n$ we say that the **tangent vector at p corresponding to x and v** is the set of all differentiable parameterizations h through p for which 0 is in the domain of h and $h(0) = p$ and $(x \circ h)'(0) = \frac{d\, x \circ h}{dt}(0) = v$.

Tangent vectors are also called **contravariant vectors** or, sometimes, simply **vectors at p**.

Parameterizations of curves through p are tested against each other by composing each with x and determining the derivative at 0. All parameterizations in $\mathcal{M}$ going in the same direction with the same speed as they pass through p *according to this test* are bundled together to form one tangent vector at p, so there is one tangent vector for each $v \in \mathbb{R}^n$. We do *not* claim two parameterizations in this tangent vector have velocity v. *But they both have that velocity when composed with x as they pass through* p at time 0.

We will denote by $[\mathbf{h}]_{\mathbf{p}}$ the set of differentiable parameterizations containing h; that is, $[h]_p$ is the tangent vector containing h. When we don't want to specify a particular member curve we might denote tangent vectors at p by upper case Latin letters, such as $\boldsymbol{X_p}$**,** $\boldsymbol{Y_p}$ or $\boldsymbol{Z_p}$, but remember that these tangent vectors correspond to sets of curves and in a calculation with tangent vectors a representative curve will often be (and can always be) extracted or produced.

Don't confuse $X_p = [h]_p$ with the derivative vector $(x \circ h)'(0) = v \in \mathbb{R}^n$. Here, v is just a parameter, the unique label (for this x) shared by all the curves in $[h]_p$.

If h is a differentiable parameterization of a curve through p and the function ${}^{\alpha}h$ defined by ${}^{\alpha}h(t) = h(t - \alpha)$ is in X_p we say $\boldsymbol{X_p}$ **is tangent to h at parameter value** $\boldsymbol{\alpha}$. This allows us to talk about tangency to curves that might pass through a point p multiple times from different directions at different times, or pass through p at some nonzero time.

Since $p = h(0)$ it is slightly redundant to refer to a tangent vector at p by $[h]_p$ and we may sometimes refer to this class of curves by $[h]$.

So the specific tangent vector consisting of all h for which $(x \circ h)'(0) = v$ is $[G^v_{x,p}]$ since $G^v_{x,p}$ is in this class.

Define the **tangent space at p** to be the collection of all these tangent vectors, each an equivalence class of parametrized curves. Exactly one of these tangent vectors is tangent to each one-to-one differentiable parameterization through p, and we have seen that the set of parameterizations passing through p at time 0 and corresponding to any particular vector in $\mathbb{R}^n$ is nonempty.

We denote the tangent space at p by the symbols $\boldsymbol{\mathcal{M}_p}$.

It remains to define vector operations on $\mathcal{M}_p$.

If r is real and $[G^v_{x,p}]$ and $[G^w_{x,p}]$ are two tangent vectors we define

$$\boldsymbol{r\,[G^v_{x,p}] = [G^{rv}_{x,p}] \quad \text{and} \quad [G^v_{x,p}] + [G^w_{x,p}] = [G^{v+w}_{x,p}].}$$

Multiplication by r introduces a "speed change" to the parameterizations. The new class consists of exactly those parameterizations which move r times faster (if r is positive) as they pass through p.

Adding two tangent vectors produces a class of parameterizations that move through p in "intermediate direction" between the two summands.

We are using the archetypal vector space $\mathbb{R}^n$ to define these operations, so the various properties required of vector operations are easily seen to be satisfied: $\mathcal{M}_p$ is actually a vector space with these operations.

For historical reasons, and by analogy with facts from Multi-Variable Calculus, we usually denote $[G^{e_i}_{x,p}]$ by $\boldsymbol{\frac{\partial}{\partial x^i}\Big|_p}$.

And because the notation is, already, cumbersome enough we will write $\boldsymbol{\frac{\partial}{\partial x^i}}$ in place of $\frac{\partial}{\partial x^i}\Big|_p$ whenever confusion will not result.

It is easy to see that $\boldsymbol{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}}$ forms a basis for $\mathcal{M}_p$ with these operations.

The **ordered basis of $\boldsymbol{\mathcal{M}_p}$ corresponding to coordinate map x** is defined to be the ordered list of tangent vectors $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$.

Normally, $[G^v_{x,p}]$ will be indicated in terms of this basis as the sum

$$v^i \frac{\partial}{\partial x^i} \;=\; v^1 \frac{\partial}{\partial x^1} + \cdots + v^n \frac{\partial}{\partial x^n}$$

where on the left we invoke the handy **Einstein summation convention.**

It is important not to forget, and we emphasize again, that the member v of $\mathbb{R}^n$ used here to select parameterizations to form a tangent vector is mentioned only in context of the coordinate map x, which connects v to the directions you can go, in $\mathcal{M}$, while passing through p.

The vector space $\mathcal{M}_p$ is isomorphic to $\mathbb{R}^n$ using coordinates in the basis corresponding to x: we send each $\frac{\partial}{\partial x^i}$ to e_i and extend by linearity. This map, $\boldsymbol{\Psi_x}$, is given by

$$\Psi_x \colon \mathcal{M}_p \to \mathbb{R}^n, \qquad \Psi_x\left(v^i \frac{\partial}{\partial x^i}\right) = v^i\, e_i.$$

So if $[h] = v^i \frac{\partial}{\partial x^i} \in \mathcal{M}_p$ then $\Psi_x\left(v^i \frac{\partial}{\partial x^i}\right)$ can be calculated as $(x \circ h)'(0) = \frac{d\, x \circ h}{dt}(0)$.

$$\Psi_x\left(v^i \frac{\partial}{\partial x^i}\right) = v^i\, e_i = (x^i \circ h)'(0)\, e_i = \begin{pmatrix} (x^1 \circ h)'(0) \\ \vdots \\ (x^n \circ h)'(0) \end{pmatrix} = (x \circ h)'(0) = \frac{d\, x \circ h}{dt}(0).$$

The coefficient v^i in $v^i \frac{\partial}{\partial x^i}$ is the derivative of the ith coordinate function of x composed with any member parameterization in $v^i \frac{\partial}{\partial x^i}$.

It remains to decide to what extent all this depends on x. The same parameterizations will be differentiable with another chart, but will the sets of parameterizations swap around their members if we choose different coordinates? Even if the sets stay the same, will vector addition and scalar multiplication be consistent with the definitions given above if new coordinates are chosen?

In section 9 the answers to these questions are found to be no, yes and yes.

8. The Cotangent Space

The cotangent space *could* be defined, simply, as the dual of the tangent space, but we will give an independent definition and recognize it as an embodiment of this dual.

Once again we choose coordinate map x around p and use it in the definition.

Recall that if $f \in \mathcal{F}_p^1$ then, by definition, $(f \circ x^{-1})'$ exists and is continuous in a neighborhood of p_x and $(f \circ x^{-1})'(p_x) = \frac{df}{dx}(p)$ is the row matrix

$$\left(\frac{\partial f}{\partial x^1}(p), \dots, \frac{\partial f}{\partial x^n}(p) \right) = \left(D_1(f \circ x^{-1})(p_x), \dots, D_n(f \circ x^{-1})(p_x) \right).$$

We define the **cotangent space at p**, denoted $\boldsymbol{\mathcal{M}_p^*}$, to be the set of equivalence classes of members of $\mathcal{F}_p^1$ where f is equivalent to g when

$$\frac{df}{dx}(p) = (f \circ x^{-1})'(p_x) = (g \circ x^{-1})'(p_x) = \frac{dg}{dx}(p).$$

Note that if the row $\sigma = (\sigma_1, \dots, \sigma_n) \in \mathbb{R}^{n*}$ is any possible value of the derivative then the real function

$$\boldsymbol{f_{x,\sigma}} : U_x \to \mathbb{R} \quad \text{given by} \quad f_{x,\sigma}(q) = \sigma_1 \, x^1(q) + \cdots + \sigma_n \, x^n(q)$$

has constant derivative σ on the whole coordinate patch U_x and is, therefore, among the class of functions determined by this derivative value at p. If we ever have need of a function in this class, this would be the simplest choice for coordinate map x.

These equivalence classes are called (synonymously) **cotangent vectors, covariant vectors, covectors or 1-forms at p.** Lower case greek letters such as $\boldsymbol{\mu_p}$ and $\boldsymbol{\phi_p}$ may be used to denote generic cotangent vectors.

Vector operations are even easier to define here than in the tangent space.

If f is in cotangent vector μ_p and g is in cotangent vector ϕ_p and r is real

$$\text{define} \quad \boldsymbol{\mu_p + r\,\phi_p} \quad \text{to be the class of} \quad f + r\,g$$

where the function $f + r\,g$ is defined on the intersection of the domains of f and g.

Linearity of the derivative operation and its local nature implies this class does not depend on the representative functions f and g used to define it, and with these operations $\mathcal{M}_p^*$ is a real vector space.

Multiplying a cotangent vector by real number r corresponds to a class of functions that is changing their values "r times faster" (if r is positive) in the vicinity of p. Adding two cotangent vectors changes the level surfaces of a representative function, which now are "intermediate" to the level surfaces of representatives of the summands near p.

When $f \in \mu_p$ we write $\mathbf{df_p} = \mu_p$ and μ_p is called the **differential of f at p**.

For each p the map $\quad \boldsymbol{d} \colon \boldsymbol{\mathcal{F}_p^1 \to \mathcal{M}_p^*} \quad$ is linear.

There is another interesting property possessed by the differential. For each p:

$$\boldsymbol{d(fg)_p = f(p)\,dg_p + g(p)\,df_p} \quad \text{for all } f, g \in \mathcal{F}_p^1.$$

The same function g could easily be in a covector at p and a covector at q for $p \neq q$. However we *cannot* have $dg_p = dg_q$ unless $p = q$. To see this let

$$\gamma(\, dg_p\,) = \bigcap_{f \in dg_p} \{\, q \mid q \text{ is in the domain of } f\,\}.$$

p must be in each of the sets in the intersection and (since $\mathcal{M}$ is Hausdorff) no other point can be in *every* domain so $\gamma(\, dg_p\,) = \{\, p\,\}$ and p can be recovered as the unique point in the domain of all the functions in *any particular covector*.

Actually, more is true. If $\mathcal{O}$ is the domain of any particular function f in covector μ_p then $\{p\}$ is the intersection of the domains of all the restrictions of f that lie in μ_p. So just the members of μ_p that are restrictions of any single member of μ_p are sufficient to "locate" μ_p as a member of $\mathcal{M}_p$.

Since p is fixed for now we may suppress mention of it when that is not ambiguous, and refer, for example, to df rather than df_p.

Of particular importance are the differentials dx^i of the coordinate functions $x^i\colon U_x \to \mathbb{R}$.

If f is a representative function in covector df it is straightforward to see that there is one and only one linear combination of these dx^i equal to df, namely

$$df = \frac{\partial f}{\partial x^1}\, dx^1 + \cdots + \frac{\partial f}{\partial x^n}\, dx^n.$$

Therefore the ordered list $\boldsymbol{dx^1, \ldots, dx^n}$ forms a basis for $\mathcal{M}_p^*$, the **ordered basis of $\mathcal{M}_p^*$ corresponding to coordinate map x** .

Define the isomorphism $\boldsymbol{\Phi_x}\colon \mathcal{M}_p^* \to \mathbb{R}^{n*}$ to be the map sending dx^i to standard dual basis vector e^i, extending by linearity.

$$\text{If } df = \sigma_i\ dx^i, \quad \Phi_x(df) = \sigma_i\, e^i = (\sigma_1, \ldots, \sigma_n) = (f \circ x^{-1})'(p_x) = \frac{df}{dx}(p).$$

The coefficient on dx^i for df is the ith partial derivative of $f \circ x^{-1}$ evaluated at p_x. It is also the derivative of $f \in df$ composed with the ith coordinate gridcurve through p for x.

We cause any $\boldsymbol{df}$ to act on any one-to-one differentiable parameterization of a curve through $\boldsymbol{p}$ with $\boldsymbol{h(\alpha) = p}$ by $\boldsymbol{df(h) = (f \circ h)'(\alpha)}$ and it is easy to check this number does not depend on the representative f for df.

But the related action of members of $\mathcal{M}_p^*$ on members of $\mathcal{M}_p$, defined precisely below, makes the critical identification of $\mathcal{M}_p^*$ as the dual of $\mathcal{M}_p$.

Suppose h is in the tangent vector $[h] = v^i \frac{\partial}{\partial x^i}$ and f is a representative of the covector $df = \sigma_i\, dx^i$.

$$\text{Define} \quad \boldsymbol{df([h]) = \sigma_i\, v^i.}$$

With this action, $\boldsymbol{df}$ is obviously a linear map from $\boldsymbol{\mathcal{M}_p}$ to $\mathbb{R}$, and $dx^i\big(\frac{\partial}{\partial x^j}\big) = \delta^i_j$. In other words, $\mathcal{M}_p^*$ "is" the dual of $\mathcal{M}_p$.

But this is a bit of a cheat: since the dimensions match we could have assigned any basis for $\mathcal{M}_p^*$ to act as the basis dual to the standard basis for $\mathcal{M}_p$ and coordinate map x. Why pick this action?

Using representatives f of covector df and h of tangent vector $[h]$ as above we have an ordinary real valued function $f \circ h$ defined on an interval. And

$$\begin{aligned}(f \circ h)'(0) = & \ (f \circ x^{-1} \circ x \circ h)'(0) = (f \circ x^{-1})'(p_x)\ (x \circ h)'(0) \\ = & \ (\sigma_1, \ldots, \sigma_n) \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \Phi_x(df)\, \Psi_x([h]) \\ = & \ \sigma_i\, v^i = \sigma_i\, dx^i\, v^j\, \frac{\partial}{\partial x^j} = df([h]).\end{aligned}$$

In other words, with this definition, causing a covector to act on a contravector is equivalent to the chain rule applied to the composition of any pair of representatives of the equivalence classes. And the calculation can be performed using the isomorphisms into $\mathbb{R}^{n*}$ and $\mathbb{R}^n$ provided by choosing the standard bases in $\mathcal{M}_p^*$ and $\mathcal{M}_p$ generated by the coordinate map x.

There is a way of thinking about, visualizing, these differentials.

Let's consider a coordinate x^i and, for specificity, x^1. This coordinate can be used to create a myriad of closely spaced "constant x^1 submanifolds" on $\mathcal{M}$ near p consisting of those points m on $\mathcal{M}$ with $x^1(m) = p_x^1 + b$ for many different small positive and negative values of b, one value for each submanifold.[9]

In a specific situation you might start with $b = 0$ and then increment b by ± 0.001 a few hundred times, thereby carving out several hundred submanifolds of U_x which may be visualized as roughly parallel (they never intersect and are close to each other and smooth and of dimension 1 less than the manifold itself) near p.

When you evaluate dx^1 on a tangent vector $v^j \frac{\partial}{\partial x^j}$ and obtain v^1, you can interpret this measurement in the following way: if you are zooming along up on $\mathcal{M}$ following a parameterization in $v^j \frac{\partial}{\partial x^j}$ then, at the instant you pass through p, the rate at which you are punching through these constant-x^1 submanifolds is v^1.

A negative v^1 would mean you are passing through submanifolds of diminishing x^1 value, while if v^1 is positive you pass through submanifolds of increasing x^1 value as time (the parameter) goes on. The magnitude of v^1 tells you how fast you can expect the b-values on these constant-x^1 submanifolds you see flashing by to increase. Of course, this will only hold for a short time interval, very near to p.

More generally, a nonzero linear combination of coordinates of the form $f = \sigma_i\, x^i$ can be used to create submanifolds up on $\mathcal{M}$ each consisting of points m with $f(m) = \sigma_i\, p_x^i + b$ for many different small values of b.

If you are passing through p on a parameterization in $v^j \frac{\partial}{\partial x^j}$ then you expect to see these constant-f submanifolds flash by at rate $\left(\sigma_i\ dx^i\right)\left(v^j \frac{\partial}{\partial x^j}\right) = \sigma_i\, v^i$.

You can think of this in two different ways: from the viewpoint of a moving eyeball on a fixed parameterization, testing different families of constant-f submanifolds or from the standpoint of a fixed f whose constant-f submanifolds are being pierced by many different travelers passing through p.

[9] By Theorem 6.6 these *are* submanifolds of dimension $n - 1$.

In the first case it is f that will vary and you are using the parameterization to measure how fast these various functions are changing in the direction and at the pace you are bound to go.

In the second case f is fixed and f "tests" parameterizations by counting how many of his (or her) constant-f submanifolds are being punctured per unit time by travelers passing through p on one parameterization or another.

These interpretations are completely clear down in $\mathbb{R}^n$. The "level hypersurfaces" there are the hyperplanes of all vectors S with $\sigma\,(S-p_x)=b$ for various small values of b and a fixed row matrix $\sigma=(\sigma_1\ \dots\ \sigma_n)$. The parameterizations can be taken to be of the form $h(t)=p_x+v\,t$, since there is a parameterization like this matching the derivative of each possible differentiable parameterization through p_x.

x^{-1} transports these surfaces and the parameterizations which pierce them (and our intuition along with) from $\mathbb{R}^n$ up to $\mathcal{M}$.

9. Change of Coordinates

Now comes the moment of truth. We need to see how all this works in a new coordinate map, and how we can translate from one allowed representation to another.

Suppose $y\colon \mathcal{U}_y\to\mathcal{R}_y$ and $x\colon \mathcal{U}_x\to\mathcal{R}_x$ are two coordinate maps around $p\in\mathcal{M}$.

Recall the coordinate gridcurves given by

$$G^{e_i}_{x,p}(t)=x^{-1}(p_x+t\,e_i) \quad \text{and} \quad G^{e_i}_{y,p}(t)=y^{-1}(p_y+t\,e_i)$$

and the two Jacobian matrices $\frac{dy}{dx}(p)=(y\circ x^{-1})'(p_x)$ and $\frac{dx}{dy}(p)=(x\circ y^{-1})'(p_y)$ which are inverse to each other.

The ijth entry of $(y\circ x^{-1})'(p_x)$ is $D_j(y^i\circ x^{-1})(p_x)=\frac{\partial y^i}{\partial x^j}(p)$ while the ijth entry of $(x\circ y^{-1})'(p_y)$ is $D_j(x^i\circ y^{-1})(p_y)=\frac{\partial x^i}{\partial y^j}(p)$. Dropping explicit reference to p, since it is fixed for the discussion, gives

$$\frac{\partial y^i}{\partial x^k}\,\frac{\partial x^k}{\partial y^j}=\delta^i_j=\frac{\partial x^i}{\partial y^k}\,\frac{\partial y^k}{\partial x^j}.$$

The jth column of $(y\circ x^{-1})'$ is the $n\times 1$ matrix $(y\circ G^{e_i}_{x,p})'(0)$. This is the derivative of the coordinate map y confined to (that is, composed with) the jth coordinate gridcurve of x.

The ith row is the $1\times n$ matrix $(y^i\circ x^{-1})'$, the derivative of the ith coordinate function $y^i\circ x^{-1}$, a real valued function of n variables.

Consider the inverse (to each other) Jacobian matrices

$$\frac{dx}{dy} = (x \circ y^{-1})'(p_y) = \big(D_j(x^i \circ y^{-1})(p_y)\big) = \left(\frac{\partial x^i}{\partial y^j}(p)\right)$$
and
$$\frac{dy}{dx} = (y \circ x^{-1})'(p_x) = \big(D_j(y^i \circ x^{-1})(p_x)\big) = \left(\frac{\partial y^i}{\partial x^j}(p)\right).$$

Suppose h is a parameterization through p with $h(0) = p$ and which has been found to be differentiable using x:

$$(x \circ h)'(0) = v \text{ for some } v \in \mathbb{R}^n \text{ so } h \text{ is a member of } v^i \frac{\partial}{\partial x^i}.$$

Then $y \circ h = y \circ x^{-1} \circ x \circ h$. By assumption, $y \circ x^{-1}$ is differentiable and so is $x \circ h$ so the chain rule implies that h is differentiable using y as the "differentiability tester." Its derivative using y is in fact the column

$$(y \circ h)'(0) = (y \circ x^{-1})'(p_x)\ (x \circ h)'(0) = \frac{dy}{dx}(p)\, \frac{d\, x \circ h}{dt}(0) = \frac{dy}{dx}\, v.$$

A symmetrical argument holds for parameterizations originally tested and found to be differentiable using y, so the same parameterizations are differentiable using either x or y.

Another glance at that derivative shows that $(y \circ h)'(0)$ is a fixed (i.e. independent of h) invertible matrix multiple of $(x \circ h)'(0)$, so the sets of parameterizations constituting tangent vectors are identical: the derivatives of the member parameterizations change, but they all change the same way, together.

So the definition of $\mathcal{M}_p$ is independent of whether you use coordinate map x or coordinate map y to define the classes of functions which comprise it.

The same argument holds for differentials. Specifically, for real valued f the function $f \circ x^{-1}$ is continuously differentiable in a neighborhood of p_x exactly when $f \circ y^{-1} = f \circ x^{-1} \circ x \circ y^{-1}$ is continuously differentiable in a neighborhood of p_y. And the derivatives of $f \circ x^{-1}$ and $g \circ x^{-1}$ are equal at p_x exactly when the derivatives of $f \circ y^{-1}$ and $g \circ y^{-1}$ are equal at p_y.

In fact if f represents the covector $df = \sigma_i\, dx^i$ then

$$(f \circ y^{-1})'(p_y) = (f \circ x^{-1})'(p_x)(x \circ y^{-1})'(p_y) = \frac{df}{dx}(p)\frac{dx}{dy}(p) = \sigma\, \frac{dx}{dy}.$$

So covectors defined using x are identical to covectors defined using y.

The question remains of whether the vector operations defined using x match those defined using y.

For covectors there is nothing to prove: vector operations are defined using ordinary addition and multiplication by constants applied to the functions chosen to represent the covectors. So if the classes of functions are the same, vector operations on them must be too.

The situation is slightly more involved for tangent vectors.

Suppose r is real and consider the combination

$$\left[G^s_{y,p}\right] + r\ \left[G^t_{y,p}\right] = \left[G^{s+r\,t}_{y,p}\right].$$

The equality in the line above is not a calculation: it is the definition of the vector operations using the coordinate map y.

We did calculate above that $s = \frac{dy}{dx}\, v$ whenever $\left[G^s_{y,p}\right] = \left[G^v_{x,p}\right]$ and by extension, using $\mathcal{B} = \frac{dy}{dx}$ (evaluated at p, of course)

$$\left[G^s_{y,p}\right] + r\,\left[G^t_{y,p}\right] = \left[G^{s+r\,t}_{y,p}\right] = \left[G^{\mathcal{B}v+r\,\mathcal{B}w}_{y,p}\right] = \left[G^{\mathcal{B}(v+r\,w)}_{y,p}\right] = \left[G^{v+r\,w}_{x,p}\right].$$

The class of functions on the far right is $\left[G^v_{x,p}\right] + r\,\left[G^w_{x,p}\right]$ calculated using the coordinate map x, and we conclude that vector operations on $\mathcal{M}_p$ agree when defined using x, y or in fact any coordinate map around p.

So with our two coordinate maps x and y we have produced 4 ordered bases:

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n} \quad \text{and} \quad \frac{\partial}{\partial y^1}, \dots, \frac{\partial}{\partial y^n} \quad \text{for} \quad \mathcal{M}_p$$

$$\text{and} \qquad dx^1, \dots, dx^n \quad \text{and} \quad dy^1, \dots, dy^n \quad \text{for} \quad \mathcal{M}^*_p.$$

The last two are the dual bases for the first two. And the Jacobian matrices $\frac{dy}{dx}(p)$ and $\frac{dx}{dy}(p)$ are the matrices of transition connecting coordinates in these bases.

This is reflected (without matrices) in the four useful facts listed below.

$$\frac{\partial}{\partial y^j} = \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial x^i} \qquad\qquad \frac{\partial}{\partial x^j} = \frac{\partial y^i}{\partial x^j}\frac{\partial}{\partial y^i}$$

$$dx^j = \frac{\partial x^j}{\partial y^i}\, dy^i \qquad\qquad dy^j = \frac{\partial y^j}{\partial x^i}\, dx^i$$

Using the isomorphisms into $\mathbb{R}^n$ and $\mathbb{R}^{n*}$ defined earlier we have

$$\begin{aligned}\Phi_y(df) &= (f \circ y^{-1})'(p_y) = (f \circ x^{-1} \circ x \circ y^{-1})'(p_y)\\ &= (f \circ x^{-1})'(p_x)\ (x \circ y^{-1})'(p_y) = \Phi_x(df)\ \frac{dx}{dy}.\end{aligned}$$

So changing ordered basis in $\mathcal{M}^*_{\mathbf{p}}$ corresponds to the chain rule implemented by matrix multiplication applied to the representatives in $\mathbb{R}^{\mathbf{n}*}$.

Now suppose we have a parameterization h in a tangent vector $[h]$.

$$\begin{aligned}\Psi_y([h]) &= (y \circ h)'(0) = (y \circ x^{-1} \circ x \circ h)'(0)\\ &= (y \circ x^{-1})'(p_x)\,(x \circ h)'(0) = \frac{dy}{dx}\ \Psi_x([h]).\end{aligned}$$

So once again, changing ordered basis in $\mathcal{M}_{\mathbf{p}}$ corresponds to the chain rule calculated by matrix multiplication applied to the representatives in $\mathbb{R}^{\mathbf{n}}$.

Thus, if df is any covariant vector and $[h]$ a contravariant vector

$$(df)\,([h]) = \Phi_y(df)\,\Psi_y([h]) = \Phi_x(df)\,\frac{dx}{dy}\,\frac{dy}{dx}\,\Psi_x([h]) = \Phi_x(df)\,\Psi_x([h]).$$

In other words, a differential evaluated on a tangent vector can be calculated in any basis and if you prefer you can do so using representative vectors and covectors in $\mathbb{R}^n$ and $\mathbb{R}^{n*}$.

Finally, we arrive at the question of whether *all possible* changes of ordered basis in $\mathcal{M}_p$ can be arrived at by a coordinate change from an arbitrary coordinate map such as x. The answer is affirmative and follows from the observation below.

Suppose W is *any* invertible $n \times n$ matrix. Any possible change of ordered basis in $\mathcal{M}_p$ must generate a matrix of transition of this kind. Let $R_y = \{\, Wu \mid u \in R_x \,\}$ and $U_y = U_x$. Then the function $y\colon U_y \to R_y$ given by $y(q) = Wx(q)$ is a coordinate map.

$$(y \circ x^{-1})' = (Wx \circ x^{-1})' = W.$$

10. Two Examples

Now that we have the basic constructions and facts in hand, let's look at a couple of examples.

First, we consider the manifold given by $\mathbb{R}$ itself. This one-dimensional situation is unique in that the identity map t, for which $t(r) = r$ for all real r, plays three roles. It is a coordinate map, and also a real valued function on the manifold and a curve in the manifold.

Other coordinate maps around a point are one-to-one smooth functions with nonzero derivative defined on an open set containing that point. The one-by-one "matrix" of transition from coordinate map x to coordinate map y at some point p is the number $(y \circ x^{-1})'(p_x)$.

Let's specify point $p = \pi/3$ in our manifold and alternative coordinate map x given by $x(r) = \tan(r)$ on the interval $(-\pi/2, \pi/2)$. So $p_x = \sqrt{3}$.

The two Jacobians are

$$(t \circ x^{-1})'(p_x) = (x^{-1})'(\sqrt{3}) = \frac{1}{4} \quad \text{and} \quad (x \circ t^{-1})'(p_t) = x'(p) = \sec^2(\pi/3) = 4.$$

The standard basis vector for $\mathbb{R}$ is the number $e_1 = 1$.

The unit pace gridcurve through $\pi/3$ is $G^1_{t,\pi/3}$ given by the time-shifted identity map

$$G^1_{t,\pi/3}(r) = \pi/3 + r \cdot 1$$

and the tangent vector corresponding to this curve is normally denoted $\left.\frac{d}{dt}\right|_{\pi/3}$, using the "straight d" rather than the curved one. This tangent vector is the set of all differentiable curves through $\pi/3$ with derivative 1 with respect to the identity coordinate map at time 0, at which time they all pass through $\pi/3$.

Any curve in a tangent vector at $\pi/3$ is a multiple of this one:

$$[h]_{\pi/3} = v \left.\frac{d}{dt}\right|_{\pi/3} \quad \text{where the number } v \text{ is } h'(0) \text{ and } h(0) = \pi/3.$$

Similarly, thinking of t as a real valued function and if f is any differentiable real valued function defined around $\pi/3$ then $df = \sigma\, dt$ where $\sigma = f'(\pi/3)$.

Causing df and $[h]$ to act on each other we have

$$df([h]) = \sigma\, v\, dt \left(\frac{d}{dt}\right) = \frac{df}{dh}\,\frac{dh}{dt} = \frac{d(f\circ h)}{dt}$$

where in the second-to-last term we have used one of the *classic* expressions of the chain rule, where $\frac{df}{dh}$ denotes $f'(h(0)) = f'(\pi/3)$.

Using better (more intelligible) notation,

$$df([h]) = f'(\pi/3)\cdot h'(0) = f'(h(0))\cdot h'(0) = (f\circ h)'(0).$$

Using the two Jacobians, we have alternative expressions for df and $[h]$ in terms of the x coordinate map.

$$\begin{aligned} df =& (f\circ x^{-1})'(\sqrt{3})\, dx = f'(\pi/3)\cdot (x^{-1})'(\sqrt{3})\, dx = \frac{\sigma}{4}\, dx \\ \text{and} \quad [h] =& (x\circ h)'(0)\,\frac{d}{dx} = x'(\pi/3)\cdot h'(0)\,\frac{d}{dx} = 4\,v\,\frac{d}{dx}. \end{aligned}$$

A small movement in the manifold generates changes in both x and in t values. But x changes faster than t, so f is changing at a smaller rate compared to x than compared to t. So df is a smaller multiple of dx than of dt.

On the other hand, $\frac{d}{dx}$ corresponds to the family of curves containing

$$G^1_{x,\sqrt{3}}(r) = \tan^{-1}(\sqrt{3} + r\cdot 1).$$

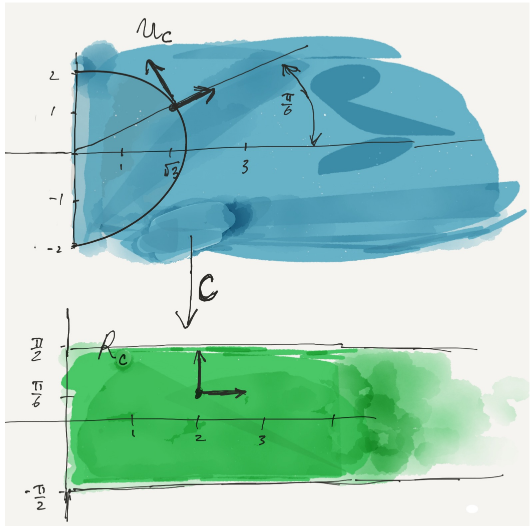
These curves are moving at a quarter of the rate that the curves in $\frac{d}{dt}$ move as they pass through $\pi/3$. Compared to these slower curves h is changing 4 times faster.

As a second example let's consider the submanifold $\mathcal{M}$ of $\mathbb{R}^2$ consisting of the half-plane $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \,\middle|\, x > 0 \right\}$ at the point $p = \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}$.

We consider two coordinate maps, the identity map and polar coordinates, each defined on all of $\mathcal{M}$. Explicitly, the identity coordinate map is given by $I\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$, which is its own chart, while polar coordinates are given by

$$C\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\begin{pmatrix} x \\ y \end{pmatrix} \\ \theta\begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2} \\ \arctan(y/x) \end{pmatrix}$$

$$\text{with chart } C^{-1}\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} x\begin{pmatrix} r \\ \theta \end{pmatrix} \\ y\begin{pmatrix} r \\ \theta \end{pmatrix} \end{pmatrix} = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}.$$



So $p_C = \begin{pmatrix} 2 \\ \pi/6 \end{pmatrix}$ and the two Jacobian matrices at $p = p_I$ and p_C are

$$(I \circ C^{-1})'(p_C) = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -r\,\sin(\theta) \\ \sin(\theta) & r\,\cos(\theta) \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & -1 \\ 1/2 & \sqrt{3} \end{pmatrix}$$

and

$$(C \circ I^{-1})'(p_I) = \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \begin{pmatrix} x/\sqrt{x^2+y^2} & y/\sqrt{x^2+y^2} \\ -y/(x^2+y^2) & x/(x^2+y^2) \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & 1/2 \\ -1/4 & \sqrt{3}/4 \end{pmatrix}.$$

Suppose given a row matrix (a, b) which represents the differential $a\,dx + b\,dy$ in the standard basis of the cotangent space using coordinate map I.

And suppose the column $\begin{pmatrix} r \\ s \end{pmatrix}$ represents the tangent vector $r\,\frac{\partial}{\partial x} + s\,\frac{\partial}{\partial y}$ in the corresponding basis of the tangent space.

Causing $a\,dx + b\,dy$ to act on $p\,\frac{\partial}{\partial x} + q\,\frac{\partial}{\partial y}$ yields the number $ap + bq$ and this number is invariant: it has physical meaning and cannot depend on the coordinate map. That means if we change to polar coordinates the coefficients must change in a coordinated way so as to preserve this invariant.

$$\text{Since } \frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} \quad \text{and} \quad \frac{\partial}{\partial y} = \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta}$$

$$\text{and} \quad dx = \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta \quad \text{and} \quad dy = \frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial \theta}d\theta$$

in our case at point p we have

$$\begin{aligned} ap+bq =&(a,b)\begin{pmatrix}p\\q\end{pmatrix} = (a,b)\begin{pmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta}\end{pmatrix}\begin{pmatrix}\frac{\partial r}{\partial x} & \frac{\partial r}{\partial y}\\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y}\end{pmatrix}\begin{pmatrix}p\\q\end{pmatrix}\\ =&(a,b)\begin{pmatrix}\sqrt{3}/2 & -1\\ 1/2 & \sqrt{3}\end{pmatrix}\begin{pmatrix}\sqrt{3}/2 & 1/2\\ -1/4 & \sqrt{3}/4\end{pmatrix}\begin{pmatrix}r\\s\end{pmatrix}\\ =&\left(\frac{\sqrt{3}}{2}a+\frac{1}{2}b,\ -a+\sqrt{3}\,b\right)\begin{pmatrix}\frac{\sqrt{3}}{2}p+\frac{1}{2}q\\ -\frac{1}{4}p+\frac{\sqrt{3}}{4}q\end{pmatrix} = (A,\,B)\begin{pmatrix}P\\Q\end{pmatrix} = AP+BQ\end{aligned}$$

where these same differentials and tangent vectors are $A\,dr + B\,d\theta$ and $P\,\frac{\partial}{\partial r} + Q\,\frac{\partial}{\partial \theta}$ in the standard bases formed from polar coordinates.

11. Point Derivations

So we now have two definitions of tangent vectors. First as a vector space whose vectors are classes of curves. Second as the dual of the space of differentials. This section provides a third way of thinking of tangent vectors.

Suppose $x\colon U_x \to R_x \subset \mathbb{R}^n$ is a coordinate map on $\mathcal{M}$ as above.

We can cause $\mathcal{M}_p$ to act on $\mathcal{F}_p^1$ according to

$$v^i\frac{\partial}{\partial x^i}(f) = v^i\frac{\partial f}{\partial x^i}(p) = v^i\,D_i(f\circ x^{-1})(p_x).$$

It is clear that if $v^i\frac{\partial f}{\partial x^i}(p) = w^i\frac{\partial f}{\partial x^i}(p)$ for all $f \in \mathcal{F}_p^\infty$ then $v^i = w^i$ for all i (since the coordinate functions x^i are in $\mathcal{F}_p^\infty$) so tangent vectors are uniquely characterized by their action on $\mathcal{F}_p^\infty$, a rather small subset of $\mathcal{F}_p^1$.

Let's denote by $W(f)$ the action of this tangent vector $v^i\frac{\partial}{\partial x^i}$ on f.

It is obvious that if r is a real number and $f,g \in \mathcal{F}_p^1$

$$W(f + r\,g) = W(f) + r\,W(g) \quad \text{and} \quad W(fg) = f(p)W(g) + g(p)W(f).$$

The last equation corresponds to the **Leibniz rule** for differentiation of ordinary functions.

A **point derivation** on $\mathcal{F}_p^k$ is any real-valued function defined on $\mathcal{F}_p^k$ and which satisfies the linearity condition and Leibniz rule.

Let $\mathcal{D}_p$ denote the set of point derivations on $\mathcal{F}_p^\infty$.

$\mathcal{D}_p$ is a vector space with vector operations defined "pointwise," i.e. for real r, point derivations W and U and any $f \in \mathcal{F}_p$

$$(W + r\,U)(f) = W(f) + r\,U(f)$$

The action of a tangent vector as described above is a point derivation on $\mathcal{F}_p^1$, so with this action (by restricting domain to $\mathcal{F}_p^\infty$) each tangent vector in the tangent space $\mathcal{M}_p$ can be identified with a member of $\mathcal{D}_p$. Distinct tangent vectors are easily seen to correspond to different members of $\mathcal{D}_p$. And the vector operations on $\mathcal{M}_p$ coincide with those in $\mathcal{D}_p$ so that $\mathcal{M}_p$ "is" a subspace of $\mathcal{D}_p$.

We will show that every point derivation on $\mathcal{F}_p^\infty$ corresponds to the action of a tangent vector on $\mathcal{F}_p^\infty$ and so the two vector spaces are isomorphic[10] providing the advertised alternative way of thinking about tangent vectors.

This will be useful: if we have a real valued linear map defined on $\mathcal{F}_p^\infty$ and can show it satisfies Leibniz rule then it will be a member of $\mathcal{M}_p$ and we can take it from there to make further deductions.

Suppose W is a generic point derivation on $\mathcal{F}_p^\infty$. We will assemble some facts about W.

Let B_r denote the ball of radius r around $p_x \in R_x$ for any r.

If $B_{3\varepsilon} \subset R_x$ for the positive number ε, then (adapting the ideas from Section 5) there is a real valued function $g\colon R_x \to [0,1]$ which is infinitely differentiable at every point of R_x and $g(q_x) = 1$ for every point q_x of B_ε and for which $g(q_x) = 0$ for every point q_x of R_x outside of $B_{2\varepsilon}$. We let $h = 1 - g$.

Now suppose $F \in \mathcal{F}_p^\infty$ and F is 0 for all points in a vicinity of p. Select ε so that $F \circ x^{-1}(q_x)$ is defined whenever $q_x \in B_{3\varepsilon}$ and $F(q) = F \circ x^{-1}(q_x) = 0$ whenever $q_x \in B_{2\varepsilon}$. Form differentiable functions g and h as above.

So F is equal to the arithmetic product of two functions $(F)\,(h \circ x)$. Applying the point derivation to this product yields

$$W(F) = W(\,(F)\,(h \circ x)\,) = F(p)W(h \circ x) + h(p_x)W(F) = 0 + 0.$$

If $G = F$ in some neighborhood of p we examine the difference $F - G$ and find that $W(F - G) = 0$ so $W(F) = W(G)$. The value of W on F is determined by the behavior of F on *any* vicinity of p, no matter how tiny. The values of F outside *any* neighborhood of p are not involved in any way.

If 1 denotes the constant function equal to 1 everywhere, we have

$$W(1) = W(1 \cdot 1) = 1 \cdot W(1) + 1 \cdot W(1) = 2\,W(1).$$

So $W(1) = 0$ and we conclude $W(F) = 0$ when F is a constant function, or even a function that is constant on any neighborhood of p.

We will show that

$$W(F) = W(x^i)\frac{\partial F}{\partial x^i} \quad \text{for all } F \in \mathcal{F}_p^\infty$$

identifying W as (the action of) the tangent vector $W(x^i)\frac{\partial}{\partial x^i}$.

Select $F \in \mathcal{F}_p^\infty$.

We suppose that $\overline{F} = F \circ x^{-1}$ is defined on some ball B with center p_x.

For any $q_x \in B$ we have

$$\begin{aligned} F(q) = \overline{F}(q_x) \; &= F(p) + \int_0^1 \frac{d}{dt}\,\overline{F}(p_x + t(q_x - p_x))\,dt \\ &= F(p) + \sum_{i=1}^n (q_x^i - p_x^i) \int_0^1 D_i(\,\overline{F}\,)(p_x + t(q_x - p_x))\,dt. \end{aligned}$$

[10] Tangent vectors are point derivations on $\mathcal{F}_p^1$. However the space of point derivations on $\mathcal{F}_p^k$ for $1 \leq k < \infty$ is known to be infinite dimensional (except in trivial cases) so $\mathcal{M}_p$ is a small subspace of these more general point derivations.

Evaluating W at this representation of F (valid only for q near our fixed p, of course, but that is all we need)

$$\begin{aligned} W(F) &= \sum_{i=1}^{n} W\Big((x^i - p_x^i) \int_0^1 D_i(\,\overline{F}\,)(p_x + t(x-p_x))\, dt \Big) \\ &= \sum_{i=1}^{n} W\big(x^i\big) \int_0^1 D_i(\,\overline{F}\,)(p_x + t(p_x-p_x))\, dt \\ &\qquad\qquad + \sum_{i=1}^{n} (p_x^i - p_x^i)\, W\Big(\int_0^1 D_i(\,\overline{F}\,)(p_x + t(x-p_x))\, dt \Big) \\ &= \sum_{i=1}^{n} W\big(x^i\big) \int_0^1 D_i(\,\overline{F}\,)(p_x)\, dt \;=\; \sum_{i=1}^{n} W\big(x^i\big)\; D_i(F\circ x^{-1})(p_x). \end{aligned}$$

This is exactly what we were trying to show, and we conclude that every point derivation on $\mathcal{F}_p^\infty$ corresponds to an an explicit tangent vector with its action restricted to the subspace $\mathcal{F}_p^\infty$ of $\mathcal{F}_p^1$:

$$W = W(x^i)\; \frac{\partial}{\partial x^i}\bigg|_p$$

The point derivation evaluated on a member F of $\mathcal{F}_p^\infty$ corresponds to this tangent vector applied to the member of $\mathcal{M}_p^*$ corresponding to F, namely $\frac{\partial F}{\partial x^j}(p)\, dx_p^j$.

$$W(F) = \left(\frac{\partial F}{\partial x^j}(p)\, dx_p^j \right) \left(W(x^i)\; \frac{\partial}{\partial x^i}\bigg|_p \right) = W(x^i) \frac{\partial F}{\partial x^i}(p).$$

So a point derivation on $\mathcal{F}_p^\infty$ is actually determined by what it does to the n coordinate functions x^i.

The argument above does use the assumption that our point derivations are on $\mathcal{F}_p^\infty$. If W had been a point derivations on $\mathcal{F}_p^k$ for positive integer k then $\int_0^1 D_i(\,\overline{F}\,)(p_x + t(x-p_x))\, dt$ is only guaranteed to be in C^{k-1} and might be outside the domain of W, causing the argument to fail.

On another note we record an obvious fact that will be used later. Suppose given four members f, g, h and k of $\mathcal{F}_p^\infty$, and suppose further that $h(p) = k(p) = 0$. If W is any point derivation and $f = g + hk$ then $W(f) = W(g)$.

In other words, if f and g differ by the product of two functions whose common value at p is 0, or any finite sum of such products, then $W(f)$ and $W(g)$ coincide.

12. The Cotangent Space as a Quotient Algebra

So far we have seen covectors at a point as equivalence classes of certain real valued functions on a manifold and recognized them as members of the dual of the tangent space at that point, defined as equivalence classes of curves in the manifold.

In the last section we saw that tangent vectors can also be thought of as point derivations and we explore the concept dual to that now.

Consider $\mathcal{F}_p^\infty$, the set of infinitely differentiable real valued functions defined an open subset of $\mathcal{M}$ containing p. Two functions f and g in $\mathcal{F}_p^\infty$ are said to be **germ-equivalent** if they coincide on some neighborhood of p. This is an equivalence relation on $\mathcal{F}_p^\infty$ and the classes are called **germs**, for historical reasons having something to do with a (slightly labored) analogy to wheat, seeds, stalks, sheaves, and so on.

If $\langle f\rangle$ is the germ of function f then every function in $\langle f\rangle$ agrees with f at p, so we can define $\langle f\rangle(p) = f(p)$.

The set of these germs form a ring with pointwise operations[11] applied to representatives and constitute a real vector space, an algebra denoted $\boldsymbol{\mathcal{G}_p}$. The set of germs whose functions are 0 at p is an ideal in this ring and a subspace, denoted $\boldsymbol{\mathcal{I}_p}$. We will define $\boldsymbol{\mathcal{I}_p^2}$ to be the ideal generated by all products fg where both f and g are in $\mathcal{I}_p$. This is the set of all finite sums of products of this form. It too is a real subspace of $\mathcal{G}_p$.

Finally, let $\boldsymbol{\mathcal{E}_p}$ denote the quotient ring $\mathcal{I}_p/\mathcal{I}_p^2$. It is a real vector space and is intrinsically defined: that is, it is built from functions deemed to be differentiable using an atlas, but otherwise coordinate maps are not involved in its formation.

Recall the discussion of Taylor polynomials from Section 2. Suppose $\langle g\rangle \in \mathcal{I}_p$ and x is a coordinate map around p. The function value $g(q) = g\circ x^{-1}(q_x)$ (possibly restricted to a small neighborhood of p or p_x) can be written as

$$\begin{aligned} g(q) =& \sum_{i=1}^n \frac{\partial g}{\partial x^i}(p)(x^i(q) - p_x^i) \\ &+ \sum_{i,j=1}^n (x^i(q) - p_x^i)(x^j(q) - p_x^j) \int_0^1 (1-u) D_{j,i}(g\circ x^{-1})(\, p_x + u(x(q) - p_x)\,)\, du. \end{aligned}$$

The germ of each summand on the last line involving integrals is in $\mathcal{I}_p^2$.

So $\langle g\rangle + \mathcal{I}_p^2 = \sum_{i=1}^n \frac{\partial g}{\partial x^i}(p)\langle x^i - p_x^i\rangle + \mathcal{I}_p^2$ and it follows from this that the members

$$\langle x^i - p_x^i\rangle + \mathcal{I}_p^2 \in \mathcal{E}_p \text{ for } i = 1, \dots, n$$

span $\mathcal{E}_p$ and therefore the dimension of $\mathcal{E}_p$ cannot exceed n.

We will show that the dual of $\mathcal{E}_p$ can be identified with the point derivations on $\mathcal{F}_p^\infty$, thereby establishing its dimension as n. $\mathcal{E}_p$ itself is then identified with the dual of the set of point derivations; that is, with the space of covectors at p.

First, suppose W is a point derivation. Define $\widetilde{W}$ on $\mathcal{E}_p$ by

$$\widetilde{W}(\langle f\rangle + \mathcal{I}_p^2) = W(f).$$

By the remarks in Section 11, $W(f)$ does not vary if you add any finite sum of products of functions which are 0 at p. And $W(f) = W(g)$ if f and g are two members of the germ of f. So $\widetilde{W}$ is well-defined on $\mathcal{E}_p$.

Linearity of $\widetilde{W}$ is easy to show, so $\widetilde{W}$ is a member of the dual of $\mathcal{E}_p$. And this operation associates $W + cU$ with $\widetilde{W} + c\,\widetilde{U}$ for real constants c and derivations W and U and so is a homomorphism from derivations to the dual of $\mathcal{E}_p$.

[11] The sum or product of two real-valued functions defined on a neighborhood of a point is defined pointwise on the intersection of their domains, which is also a neighborhood of that point.

On the other hand, suppose α is a member of the dual of $\mathcal{E}_p$.

Define $\overline{\alpha}$ on $\mathcal{F}_p^\infty$ by

$$\overline{\alpha}(f) = \alpha(\ \langle f - f(p)\rangle + \mathcal{I}_p^2\ ).$$

It is easy to see that $\overline{\alpha}$ is linear on $\mathcal{F}_p^\infty$. And if f and g are in $\mathcal{F}_p^\infty$

$$\begin{aligned}\overline{\alpha}(fg) &= \alpha(\ \langle fg - f(p)g(p)\rangle + \mathcal{I}_p^2\ )\\ &= \alpha(\ \langle (f-f(p))(g-g(p)) + f(p)(g-g(p)) + (f-f(p))g(p)\rangle + \mathcal{I}_p^2\ )\\ &= \alpha(\ \langle f(p)(g-g(p))\rangle + \mathcal{I}_p^2\ ) + \alpha(\ \langle (f-f(p))g(p)\rangle + \mathcal{I}_p^2\ )\\ &= f(p)\,\alpha(\ \langle g-g(p)\rangle + \mathcal{I}_p^2\ ) + g(p)\,\alpha(\ \langle f-f(p)\rangle + \mathcal{I}_p^2\ )\\ &= f(p)\,\overline{\alpha}(g) + g(p)\,\overline{\alpha}(f).\end{aligned}$$

So $\overline{\alpha}$ is a derivation. And if W is any derivation and $f \in \mathcal{F}_p^\infty$

$$\overline{\widetilde{W}}(f) = \widetilde{W}(\ \langle f - f(p)\rangle + \mathcal{I}_p^2\ ) = W(f - f(p)) = W(f).$$

So the dual of $\mathcal{E}_p$ (and hence $\mathcal{E}_p$ itself) has dimension at least n, and coupled with our earlier observation its dimension is exactly n. These two associations are inverse (to each other) isomorphisms.

13. The Tangent and Cotangent Manifolds

Let's suppose given a smooth n-manifold $\mathcal{M}$.

The **tangent bundle for $\mathcal{M}$** is a set endowed with several types of structure. The underlying set is

$$\boldsymbol{\mathcal{T}(\mathcal{M})} = \bigcup_{p\in\mathcal{M}} \mathcal{M}_p.$$

We define $\boldsymbol{\pi}\colon \mathcal{T}(\mathcal{M}) \to \mathcal{M}$ by $\pi([h]_p) = p$. In this context $\pi^{-1}(p) = \mathcal{M}_p$ is called the **fiber over p** and as we have seen each fiber is isomorphic to $\mathbb{R}^n$.

For each coordinate map $x\colon U_x \to R_x \subset \mathbb{R}^n$ we identify $\mathcal{T}(U_x)$ with $\bigcup_{p\in U_x}\mathcal{M}_p$ and define the map

$$\underset{\sim}{\boldsymbol{x}} : \mathcal{T}(U_x) \to R_x \times \mathbb{R}^n \subset \mathbb{R}^n \times \mathbb{R}^n$$

by $\underset{\sim}{x}(X_p) = (\ x(p), \Psi_x(X_p)\ )$ or, using coordinates,

$$\underset{\sim}{x}\left(v^i \left.\frac{\partial}{\partial x^i}\right|_p \right) = \left(x(p), \Psi_x\left(v^i \left.\frac{\partial}{\partial x^i}\right|_p \right)\right) = (\ p_x,\ v)$$

where v is a member of $\mathbb{R}^n$, a column with coordinates $v^1, \ldots, v^n$.

Giving $\mathbb{R}^n \times \mathbb{R}^n$ its natural topology, we define the topology on $\mathcal{T}(\mathcal{M})$ to be the weakest topology for which all these $\underset{\sim}{x}$ are continuous for all x in any atlas generating the differentiable structure for $\mathcal{M}$.

We then make $\mathcal{T}(\mathcal{M})$ into a manifold, the **tangent manifold**, using the set of these $\underset{\sim}{x}$ as the atlas to generate the differentiable structure.[12]

---

[12]Strictly speaking, these are not coordinate maps. Their codomain is $\mathbb{R}^n \times \mathbb{R}^n$, not $\mathbb{R}^{2n}$. Make the obvious identification to deal with this technicality.

If $\mathcal{K}$ is a k-dimensional submanifold containing p every differentiable parameterization of a curve through p in $\mathcal{K}$ is *also* a differentiable parameterization of a curve through p in $\mathcal{M}$ so the equivalence classes that comprise the tangent vectors in $\mathcal{K}_p$ are each subsets of corresponding tangent vectors in $\mathcal{M}_p$. In this context we will identify these tangent vectors, and so regard $\mathcal{K}_p$ as a vector subspace of $\mathcal{M}_p$.

Each $\mathcal{T}(\mathcal{K})$ is a submanifold of $\mathcal{T}(\mathcal{M})$, and in particular each $\mathcal{T}(U_x)$ is an open submanifold of $\mathcal{T}(\mathcal{M})$.

Sometimes $\mathcal{T}(\mathcal{M})$ is diffeomorphic to $\mathcal{M}\times\mathbb{R}^n$, in which case the tangent manifold is called **trivial.** Lie groups are important examples of manifolds with trivial tangent manifold, but there are others.

Following the same procedure, we define the **cotangent bundle for $\mathcal{M}$**, as a set, by

$$\boldsymbol{\mathcal{T}^*(\mathcal{M})} = \bigcup_{p\in\mathcal{M}} \mathcal{M}_p^*.$$

Again, there is a natural projection π given by $\boldsymbol{\pi}\colon \mathcal{T}^*(\mathcal{M}) \to \mathcal{M}$ by $\pi(\, df_p\,) = p$.

$\pi^{-1}(p) = \mathcal{M}_p^*$ is called the **fiber over p** for this cotangent bundle. Here, fibers are explicitly seen to be isomorphic to $\mathbb{R}^{n*}$.

For each coordinate map $x\colon U_x \to R_x \subset \mathbb{R}^n$ we identify $\mathcal{T}^*(U_x)$ with $\bigcup_{p\in U_x}\mathcal{M}_p^*$ and define

$$\widetilde{\boldsymbol{x}}\colon \mathcal{T}^*(U_x) \to R_x\times\mathbb{R}^{n*} \subset \mathbb{R}^n\times\mathbb{R}^{n*}$$

by $\widetilde{x}(\tau_p) = (\, x(p), \Phi(\tau_p)\,)$ and in coordinates this is

$$\widetilde{x}\left(\theta_i\, dx_p^i\right) = \left(x(p), \Phi_x\left(\theta_i\, dx_p^i\right)\right) = (\, p_x,\, \theta).$$

Giving $\mathbb{R}^n\times\mathbb{R}^{n*}$ its natural topology, we define the topology on $\mathcal{T}^*(\mathcal{M})$ to be the weakest topology for which all these $\widetilde{x}$ are continuous for all x in any atlas generating the differentiable structure for $\mathcal{M}$.

$\mathcal{T}^*(\mathcal{M})$ is a manifold, the **cotangent manifold**, using the set of these $\widetilde{x}$ as a generating atlas.[13]

Each $\mathcal{T}^*(U_x)$ is, as with the tangent space, an open submanifold of $\mathcal{T}^*(\mathcal{M})$. However if $\mathcal{K}$ is a submanifold of $\mathcal{M}$ of dimension *lower* than $\mathcal{M}$ there will be many incompatible ways to extend a smooth function $f\colon \mathcal{N}\to\mathbb{R}$ to all of $\mathcal{M}$ so $\mathcal{T}^*(\mathcal{N})$ cannot be automatically regarded as a submanifold of $\mathcal{T}^*(\mathcal{M})$.

As a final comment, we note that the atlases used to generate the differentiable structures on $\mathcal{T}(\mathcal{M})$ and $\mathcal{T}^*(\mathcal{M})$ only contain coordinate maps of a special type. In each case the first n coordinates are used to identify a member p of $\mathcal{M}$ while the last n coordinates specify the member of $\mathcal{M}_p$ or $\mathcal{M}_p^*$ *using a special basis* which is determined by the first n coordinates. No "mixing" of coordinates from these two groups can occur and if there is a reason to use a different basis for the last n coordinates than the natural one produced by the first n coordinates the map using these will not appear among those in the generating atlas.

You might wonder why anyone would choose to describe a coordinate patch in the tangent or cotangent manifold using one of these unusual coordinate maps. One

[13]The codomain is $\mathbb{R}^n\times\mathbb{R}^{n*}$, not $\mathbb{R}^{2n}$. We make the obvious identification and declare these $\widetilde{x}$ to be coordinate maps.

answer comes from the universe of **Hamiltonian and Lagrangian Mechanics**, where these coordinates are most helpful in describing "constants of the motion" or periodicities and other important features of a physical system as it evolves from one state to another. The language of Differential Geometry (in its older form originally, but in its fully modern form in most recent treatments) is used to clarify and extend those subjects. The transformation from simpler coordinate maps to one of the more useful ones is implemented through a so-called **canonical transformation** of coordinates.

14. One Final Representation of Tangent and Cotangent Spaces

The representation of tangent vectors as equivalence classes of curves in the manifold and covectors as equivalence classes of real functions defined on the manifold has the advantage that these definitions are intrinsic. However they *do* depend on the atlas. The functions in the equivalence classes are functions deemed to be differentiable by that atlas. Unless you have some *other* way of knowing which real-valued functions and curves are to be differentiable, coordinate maps are involved from the very beginning.

A similar point can be made for the representation of tangent vectors as members of $\mathcal{D}_p$, the point derivations on $\mathcal{F}_p^\infty$, or covectors as members of the quotient ring $\mathcal{E}_p = \mathcal{I}_p / \mathcal{I}_p^2$.

Given that, the following definition of tangent vector has some advantages, even if it does include specific reference to charts. You really can't get away from them.

For n-manifold $\mathcal{M}$ with differentiable structure $\mathcal{A}$ and $p \in \mathcal{M}$ let

$$\overline{\mathcal{S}}_p = \{ (x, p, v) \mid x \in \mathcal{A} \text{ with } p \in U_x \text{ and } v \in \mathbb{R}^n \}.$$

Define equivalence relation $\sim$ on $\overline{\mathcal{S}}_p$ by

$$(x, p, v) \sim (y, p, w) \qquad \Longleftrightarrow \; (y \circ x^{-1})'(p_x)\, v = w.$$

Thus, in case of equivalence, $w^i = \frac{\partial y^i}{\partial x^j}(p)\, v^j$ and $v^i = \frac{\partial x^i}{\partial y^j}(p)\, w^j$.

Using Jacobian matrix $\frac{dy}{dx} = \left(\frac{\partial y^i}{\partial x^j}(p) \right) = (y \circ x^{-1})'(p_x)$ and its inverse we have

$$w = \frac{dy}{dx}\, v \quad \text{and} \quad v = \frac{dx}{dy}\, w.$$

Define $[x, v]_p$ to be the equivalence class of (x, p, v) and let $\mathcal{S}_p$ be the set of equivalence classes on $\overline{\mathcal{S}}_p$. We make $\mathcal{S}_p$ into a vector space as follows.

Define $[x, v]_p \oplus_x [x, w]_p = [x, v + w]_p$ and $r \odot_x [x, w]_p = [x, r\, w]_p$.

Suppose $[x, v]_p = [y, w]_p$ and $[x, z]_p = [y, u]_p$. Then

$$[x, v]_p \oplus_x [x, z]_p = [x, v + z]_p = \left[y, \frac{dy}{dx}(v + z) \right]_p$$

$$= \left[y, \frac{dy}{dx}\, v \right]_p \oplus_y \left[y, \frac{dy}{dx}\, z \right]_p = [y, w]_p \oplus_y [y, u]_p$$

so this addition does not depend on the coordinate map used to define it. Similarly, the scalar multiplication is independent of chart. And these operations clearly satisfy the requirements for vector space operations, making $\mathcal{S}_p$ with these operations, henceforth denoted $+$ and $\cdot$, into a vector space.

Now define $\mathcal{S}(\mathcal{M}) = \bigcup_{p\in\mathcal{M}} \mathcal{S}_p$ and for each coordinate map $x\colon U_x \to R_x$ define $\underline{x}\colon \mathcal{S}(U_x) \to R_x \times \mathbb{R}^n$ by $\underline{x}([x,v]_p) = (p_x, v)$ and give $\mathcal{S}(\mathcal{M})$ the weakest topology making all these $\underline{x}$ continuous.

$\mathcal{S}(\mathcal{M})$ is a differentiable $2n$-manifold[14] with this topology and differentiable structure generated by these $\underline{x}$, and with this structure $\mathcal{S}(\mathcal{M})$ is obviously diffeomorphic to $\mathcal{T}(\mathcal{M})$.

The diffeomorphism associates each $X_p = v^i \frac{\partial}{\partial x^i}\big|_p = [c]_p$ to

$$[x, \Psi_x(X_p)]_p = [x, (x\circ c)'(0)]_p = \left[x, \frac{d\, x\circ c}{dt}(0)\right]_p = [x,v]_p.$$

There are advantages to using $\mathcal{S}(\mathcal{M})$, mostly associated with the effort required to write down one of its members, and the distinction between $\mathcal{S}(\mathcal{M})$ and $\mathcal{T}(\mathcal{M})$ can be regarded as, essentially, notational. So we will feel free to refer to a member of $\mathcal{T}(\mathcal{M})$ using the member of $\mathcal{S}(\mathcal{M})$ with which it is identified.

In a similar way, we define

$$\overline{\mathcal{S}}^*_p = \{\, (x,p,\sigma) \mid x \in \mathcal{A} \text{ with } p \in U_x \text{ and } \sigma \in \mathbb{R}^{n*} \,\}.$$

Define equivalence relation $\sim$ on $\overline{\mathcal{S}}^*_p$ by

$$(x,p,\sigma) \sim (y,p,\tau) \quad\iff \tau = \sigma\,(x\circ y^{-1})'(p_y).$$

Thus, in case of equivalence, $\tau_j = \sigma_i \frac{\partial x^i}{\partial y^j}(p)$ and $\sigma_j = \tau_i \frac{\partial y^i}{\partial x^j}(p)$ or, using Jacobians,

$$\tau = \sigma\,\frac{dx}{dy} \quad\text{and}\quad \sigma = \tau\,\frac{dy}{dx}.$$

Define $[x,\sigma]_p$ to be the equivalence class of (x,p,σ) and let $\mathcal{S}^*_p$ be the set of these equivalence classes on $\overline{\mathcal{S}}^*_p$.

We make $\mathcal{S}^*_p$ into a vector space just as we did for $\mathcal{S}_p$.

Define $[x,\sigma]_p + [x,\mu]_p = [x, \sigma+\mu]_p$ and $r\,[x,\sigma]_p = [x, r\,\sigma]_p$.

By very similar argument to that given above, these operations are independent of chart and make $\mathcal{S}^*_p$ into a vector space.

Now define $\mathcal{S}^*(\mathcal{M}) = \bigcup_{p\in\mathcal{M}} \mathcal{S}^*_p$ and for each coordinate map $x\colon U_x \to R_x$ define $\overline{x}\colon \mathcal{S}(U_x) \to R_x \times \mathbb{R}^{n*}$ by $\overline{x}([x,\sigma]_p) = (p_x,\sigma)$ and give $\mathcal{S}^*(\mathcal{M})$ the weakest topology making all these $\overline{x}$ continuous.

As before, $\mathcal{S}^*(\mathcal{M})$ is a differentiable $2n$-manifold[15] with this topology and differentiable structure generated by these $\overline{x}$.

---

[14] Here we identify $\mathbb{R}^n \times \mathbb{R}^n$ with $\mathbb{R}^{2n}$.

[15] Actually, $\overline{x}$ is not a coordinate map. The map sending $[x,\sigma]_p$ to (p_x, σ^t), where the transpose on the second factor maps $\mathbb{R}^{n*}$ to $\mathbb{R}^n$, is the one we really want.

$\mathcal{S}^*(\mathcal{M})$ is, again, diffeomorphic to $\mathcal{T}^*(\mathcal{M})$ with diffeomorphism associating each $\tau_p = df_p = \sigma_i\, dx_p^i$ with

$$[x, \Phi_x(\tau_p)]_p = [x, (f \circ x^{-1})'(p_x)]_p = \left[x, \frac{df}{dx}(p)\right]_p = [x, \sigma]_p.$$

When convenient, we may refer to a member of $\mathcal{T}^*(\mathcal{M})$ using the associated member of $\mathcal{S}^*(\mathcal{M})$.

$\mathcal{S}(\mathcal{M})$ and $\mathcal{S}^*(\mathcal{M})$ may be the cleanest representations of $\mathcal{T}(\mathcal{M})$ and $\mathcal{T}^*(\mathcal{M})$, respectively. However $v \in \mathbb{R}^n$ and covector $\sigma \in \mathbb{R}^{n*}$ in this representation are somewhat removed from curves in the manifold representing the actual motion of particles or real valued functions on a manifold, representing something like temperature.

It is not uncommon to see discussion begin with curves or real functions as in $\mathcal{T}(\mathcal{M})$ and $\mathcal{T}^*(\mathcal{M})$ followed by calculations using v and σ from $\mathcal{S}(\mathcal{M})$ and $\mathcal{S}^*(\mathcal{M})$, with coordinate map x and point p often left out as "understood."

Some authors are dedicated to a single representation and refuse to use techniques better adapted to another. This is inefficient, since all of the aspects of tangent and cotangent vectors which we have discussed are actually used.

On the other hand, hopping around among representations, where you pick the representation that makes the calculation you are trying to do as easy as possible, can contribute to a certain amount of confusion.

As a case in point, we have defined actions of tangent vectors on three different sets: $\mathcal{F}_p^\infty$ and two versions of the cotangent space.

Suppose $X_p = [c]_p$ is a tangent vector and $\sigma_p = df_p$ for $f \in \mathcal{F}_p^\infty$ is in $\mathcal{M}_p^*$.

We have

$$\begin{aligned}\sigma_p\, X_p &= df_p\, X_p = X_p(\ \langle f - f(p)\rangle + \mathcal{I}_p^2\ ) \\ &= X_p(f) = (f \circ c)'(0).\end{aligned}$$

$X_p(f) = df_p\, X_p$ looks a bit odd, but these (equal) numbers represent the action of X_p on $\mathcal{F}_p^\infty$ and the equivalence classes that comprise the cotangent space, respectively.

These actions, though related, *have different properties.* For instance the action of X_p on $\mathcal{F}_p^\infty$ has an infinite dimensional kernel while the action on $\mathcal{M}_p^*$ does not.

And σ_p acts on individual one-to-one smooth curves b with $b(\alpha) = p$ as well as equivalence classes $[c]_p$ of curves by the same general formula:

$$\sigma_p(b) = (f \circ b)'(\alpha) \quad \text{and} \quad \sigma_p\, [c]_p = (f \circ c)'(0).$$

These two actions are *not* the same, they have different properties. For instance the first action on the set of one-to-one curves through p is not linear because that set of curves is not a vector space and this failure is irreparable. The second action on the *classes* of curves *is* linear, and its properties make the cotangent space the dual of the tangent space.

The reader is hereby warned: ***From this point on we will not distinguish among these various representations unless strictly necessary, using whichever of their essential features and applying whichever action seems most convenient.***

The next section consists of some tables to help keep it all straight.

15. The Tables to Help Keep It All Straight

In the shortest form, our assembled facts give:

$v = (x \circ c)'(0) = (x^i \circ c)'(0)\, e_i = \frac{d(x^i \circ c)}{dt}(0)\, e_i$	$\sigma = (g \circ x^{-1})'(p_x) = \frac{\partial g}{\partial x^i}(p)\, e^i$
$[x, v]_p \longleftrightarrow [c]_p \longleftrightarrow v^i \left. \frac{\partial}{\partial x^i} \right\|_p$	$[x, \sigma]_p \longleftrightarrow dg_p \longleftrightarrow \sigma_i\, dx^i_p.$

With more detail, we have the following.

$\mathcal{M}_p$	$[c]_p$	$[c]_p$ is the equivalence class of those differentiable curves c through p for which $v = \frac{d\, x \circ c}{dt}(0) = (x \circ c)'(0).$ $G^v_{x,p}$ given by $G^v_{x,p}(t) = x^{-1}(p_x + t\, v)$ is a convenient curve in this tangent vector. Vector operations are given by $[G^v_{x,p}]_p + r\, [G^w_{x,p}]_p = [G^{v+r\, w}_{x,p}]_p.$ Here, for each i, $\left. \frac{\partial}{\partial x^i} \right\|_p$ is defined as $[x^i]_p$ so $[c]_p = v^i \left. \frac{\partial}{\partial x^i} \right\|_p.$
$\mathcal{D}_p$	$v^i \left. \frac{\partial}{\partial x^i} \right\|_p$	Here the symbols $\left. \frac{\partial}{\partial x^i} \right\|_p$ denote point derivations on $\mathcal{F}^\infty_p$. These are directional derivatives acting on real-valued functions defined around p. $\left. \frac{\partial}{\partial x^i} \right\|_p$ is the point derivation that gives a rate-of-change 1 when acting on the coordinate x^i. Generally $\left. \frac{\partial}{\partial x^i} \right\|_p (f) = D_i(f \circ x^{-1})(p_x) = \frac{\partial f}{\partial x^i}(p).$ And if c is a curve in the previous representation and $f \in \mathcal{F}^\infty_p$ then $[c]_p = v^i \left. \frac{\partial}{\partial x^i} \right\|_p$ and $\left(v^i \left. \frac{\partial}{\partial x^i} \right\|_p \right)(f) = (f \circ c)'(0) = \frac{d\, f \circ c}{dt}(0).$
$\mathcal{S}_p$	$[x, v]_p$	The interpretation of these triples is given above. The notation is stripped here to bare essentials. Vector operations are given by: $[x, v]_p + r\, [x, w]_p = [x, v + r\, w]_p.$

$\mathcal{M}_p^*$	$df_p = \sigma_i\, dx_p^i$ $= \frac{\partial f}{\partial x^i}(p)\, dx_p^i$	df_p is an equivalence class of real-valued functions g defined around p for which $\sigma = (g \circ x^{-1})'(p_x)$ and the function f is a representative of this class. This means $\sigma_i = D_i(g \circ x^{-1})(p_x) = \frac{\partial g}{\partial x^i}(p)$. The simplest representative is $\sigma_i\, x^i$. Vector operations are given by $df_p + r\, dg_p = d(f + r\, g)_p.$
$\mathcal{E}_p = \mathcal{I}_p/\mathcal{I}_p^2$	$\sum_{i=1}^{n} \sigma_i \langle x^i - p_x^i \rangle + \mathcal{I}_p^2$	$\sigma_i \langle x^i - p_x^i \rangle$ is the germ of a real-valued function that is 0 at p. The sum is 0 there too and hence in $\mathcal{I}_p$. We learned that every smooth real-valued function g for which $g(p) = 0$ and $(g \circ x^{-1})'(p_x) = \sigma$ differs from this one by a member of $\mathcal{I}_p^2$.
$\mathcal{S}_p^*$	$[x, \sigma]_p$	Vector operations are given by: $[x, \sigma]_p + r\, [x, \tau]_p = [x, \sigma + r\, \tau]_p.$

Tangent vectors and covectors act on each other in coordinate-independent form using representatives of their defining function classes. But for changing coordinate maps one either uses the $\mathcal{S}_p$ and $\mathcal{S}_p^*$ forms or gyrations that are equivalent to that.

	Act on each other by ...	Go from x coordinates σ and v to y coordinates τ and w by ...
$\mathcal{M}_p$ and $\mathcal{M}_p^*$	$\sigma_j\, dx_p^j\, v^i \left.\frac{\partial}{\partial x^i}\right\|_p = \sigma_i\, v^i = \sigma\, v$ or $\quad df_p\, [c]_p = (f \circ c)'(0)$	$\tau = \sigma \left(\frac{\partial x^i}{\partial y^j}(p) \right) = \sigma\, \frac{dx}{dy}$ $w = \left(\frac{\partial y^i}{\partial x^j}(p) \right) v = \frac{dy}{dx}\, v$
$\mathcal{D}_p$ and $\mathcal{E}_p$	$v^i \left.\frac{\partial}{\partial x^i}\right\|_p \left(\sum_{i=1}^{n} \sigma_i \langle x^i - p_x^i \rangle + \mathcal{I}_p^2 \right) = \sigma\, v$ or $\quad [c]_p(\langle g \rangle + \mathcal{I}_p^2) = (g \circ c)'(0)$	...
$\mathcal{S}_p$ and $\mathcal{S}_p^*$	$[x, \sigma]_p [x, v]_p = \sigma\, v$	...

16. A Closer Look At Zero

This short section is unusual for the early part of these notes because it uses *second* derivatives. We point out the intriguing idea that there is a "copy" of the whole tangent space at p *inside each zero tangent vector.*

Suppose c is any curve in the zero tangent vector at p. So $0_p = [c]_p = [p]_p \in \mathcal{M}_p$, where in the last term we abuse notation to double-use a symbol, denoting both the point itself and the constant "curve" at that point by p.

So $c(0) = p$ and $(x \circ c)'(0) = 0$ for one, and hence every, coordinate map x around p. We have $[c]_p = [x, 0]_p$ in any coordinates.

The curve $c(t) = x^{-1}\left(p_x + \frac{t^2}{2}\, v\right)$ for t near 0 and vector $v \in \mathbb{R}^n$ is an example of a curve like this. In this case $(x \circ c)''(0) = v$.

Because $(x \circ c)'(0) = 0$ the curve c has an unusual property: its second derivative changes in new coordinates in the same way the first derivatives change with more general curves through p. If y is a second coordinate map around p

$$\begin{aligned}(y \circ c)'(t) &= (y \circ x^{-1} \circ x \circ c)'(t)\\ &= (y \circ x^{-1})'(c(t)_x)\,(x \circ c)'(t) = A(t)\,(x \circ c)'(t)\end{aligned}$$

The last term is a matrix product of two functions of time. The product rule has

$$(y \circ c)''(0) = A'(0)\,(x \circ c)'(0) + A(0)\,(x \circ c)''(0) = (y \circ x^{-1})'(p_x)\,(x \circ c)''(0).$$

We conclude that if $v = (x \circ c)''(0)$ and $w = (y \circ c)''(0)$ then $[x, v]_p = [y, w]_p$.

Define the function $Y_c \colon \mathcal{F}_p^\infty \to \mathcal{R}$ by

$$Y_c(f) = (f \circ c)''(0).$$

Y_c is obviously linear. In coordinates this is calculated in two steps as

$$\begin{aligned}(f \circ c)'(t) &= (f \circ x^{-1} \circ x \circ c)'(t)\\ &= (f \circ x^{-1})'(c(t)_x)\,(x \circ c)'(t)\end{aligned}$$

for row matrix function $(f \circ x^{-1})'(c(t)_x)$ of time. Calculating the second derivative by the product rule and using $(x \circ c)'(0) = 0$ gives

$$\begin{aligned}Y_c(f) &= (f \circ c)''(0) = (f \circ x^{-1})'(p_x)\,(x \circ c)''(0)\\ &= (f \circ x^{-1})'(p_x)\,(x \circ y^{-1})'(p_y)\,(y \circ x^{-1})'(p_x)\,(x \circ c)''(0)\\ &= (f \circ y^{-1})'(p_y)\,(y \circ c)''(0).\end{aligned}$$

This calculation illustrates several things.

For instance, if $g \in df_p$ then $Y_c(g) = Y_c(f)$.

And if $df_p = [x, \sigma]_p$ and $(x \circ c)''(0) = v$ then

$$Y_c(f) = \sigma\, v = [x, \sigma]_p\, [x, v]_p.$$

In other words, $Y_c = [x, v]_p$.

Every member of $\mathcal{M}_p$ corresponds to Y_c for a class of curves $c \in 0_p = [x, 0]_p$, and these classes exhaust 0_p.

17. The Tangent Map (Pushforward) on Tangent Spaces

We are going to think about how to transport (information about) tangent and cotangent vectors from one manifold to another; we will produces functions which implement this association between the two tangent spaces and also, in the next section, between the two cotangent spaces.

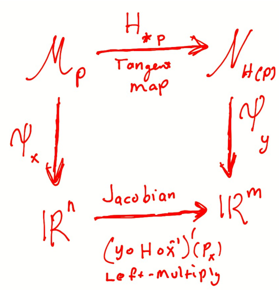
Suppose $H\colon \mathcal{M} \to \mathcal{N}$ is a smooth map between manifolds and x and y are coordinate maps around $p \in \mathcal{M}$ and $H(p) \in \mathcal{N}$, respectively.

Select tangent vector $X_p = [x, v]_p \in \mathcal{M}_p$.

Define $\boldsymbol{H_{*p}\colon \mathcal{M}_p \to \mathcal{N}_{H(p)}}$ by

$$H_{*p}([x,v]_p) = [y, (y \circ H \circ x^{-1})'(p_x)\, v]_{H(p)} = \left[y, \frac{d\, y \circ H}{dx}\, v\right]_{H(p)} = [y, w]_{H(p)}.$$

***In words, you left-multiply* v *by the Jacobian of* H *to get* w.**



This map is interpreted as the best linear approximation to H near p, applied to $\mathcal{M}_p$ and $\mathcal{N}_{H(p)}$, which are themselves the best linear approximations to $\mathcal{M}$ and $\mathcal{N}$ near p and $H(p)$, respectively.

It is, of course, necessary to verify that this map does not depend on charts and we check that now.

Suppose $\overline{y}$ and $\overline{x}$ are also coordinate maps around $H(p)$ and p, respectively.

Thus $[x, v]_p = \left[\overline{x}, \frac{d\overline{x}}{dx}\, v\right]_p = [\overline{x}, \overline{v}]_p$.

Applying the chain rule gives

$$\left[\overline{y}, \frac{d\,\overline{y} \circ H}{d\overline{x}}\, \overline{v}\right]_{H(p)} = \left[\overline{y}, \frac{d\,\overline{y} \circ H}{d\overline{x}}\, \frac{d\overline{x}}{dx}\, v\right]_{H(p)}$$
$$= \left[y, \frac{dy}{d\overline{y}}\, \frac{d\,\overline{y} \circ H}{d\overline{x}}\, \frac{d\overline{x}}{dx}\, v\right]_{H(p)} = \left[y, \frac{d\, y \circ H}{dx}\, v\right]_{H(p)}.$$

So you produce the same tangent vector in $\mathcal{N}_{H(p)}$ whichever coordinate maps are used to calculate H_{*p}.

For each $p \in \mathcal{M}$ this map is linear, and $H_{*p}(\mathcal{M}_p)$ is a subspace of $\mathcal{N}_{H(p)}$ whose dimension is the rank of H at p.

We call H_{*p} the **pushforward or tangent map of H at p**. Using the active verb, H_{*p} is said to **push** a tangent vector into $\mathcal{N}_{H(p)}$.

H_{*p} takes the class of a curve h through p to the class of a curve through $H(p)$ in the most direct way possible, by composition to form $H \circ h$.

If $[h]_p \in \mathcal{M}_p$ then $H \circ h$ is indeed a differentiable curve and at $H(p)$ at time 0 and therefore a member of a tangent vector in $\mathcal{N}_{H(p)}$. We have a fixed value for $(x \circ g)'(0)$ whenever $g \in [h]_p$ so for coordinate maps x and y

$$\begin{aligned}(y \circ H \circ g)'(0) &= (y \circ H \circ x^{-1})'(p_x)\,(x \circ g)'(0)\\ &= (y \circ H \circ x^{-1})'(p_x)\,(x \circ h)'(0) = (y \circ H \circ h)'(0).\end{aligned}$$

Therefore this tangent vector is not dependent on choice of representative h.

For every $X_p \in \mathcal{M}_p$ the tangent map for smooth $H\colon \mathcal{M} \to \mathcal{N}$ at p

$\boldsymbol{H_{*p}\colon \mathcal{M}_p \to \mathcal{N}_{H(p)}}$ may be given by $H_{*p}(X_p) = H_{*p}(\,[h]_p\,) = [H \circ h]_{H(p)}$

for any curve $h \in X_p$.

If we examine Theorem 6.3 we see how to use special coordinate maps x and y to "straighten out" H. When H has rank j at p there are coordinate maps x and y for which

$$y \circ H \circ x^{-1} \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} a^1 \\ \vdots \\ a^j \\ \psi^{j+1}(a) \\ \vdots \\ \psi^k(a) \end{pmatrix} \quad \text{for all} \quad a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \quad \text{near} \quad p_x = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If H has constant rank in a neighborhood of p the ψ^i may be chosen to be identically zero, or are absent if $j = k$.

In any case, the Jacobian $(y \circ H \circ x^{-1})'$ has the $j \times j$ identity matrix in the upper left block in a neighborhood of p_x.

If $v \in \mathbb{R}^n$ define curve b_v by $b_v(t) = H \circ x^{-1}(tv)$ for t in some (possibly) small interval around 0. ***This is a smooth curve entirely in*** $\boldsymbol{H(U_x) \subset H(\mathcal{M})}$.

$$[b_v]_{H(p)} = [y, (y \circ b_v)'(0)]_{H(p)} = [y, (y \circ H \circ x^{-1})'(p_x)\,v]_{H(p)}.$$

The first j coordinate gridcurves formed from x are, when composed with H, taken to curves corresponding to j linearly independent tangent vectors which therefore span the range of H_{*p}.

In case H has constant rank in a neighborhood of p we have a cleaner result.

17.1. ***Theorem.*** *If H has constant rank j in a neighborhood of p we can pick coordinate maps x and y so*

$$H_{*p}\left(\,\left[\,G^{e_i}_{x,p}\,\right]_p\right) = \left[\,G^{e_i}_{y,H(p)}\,\right]_{H(p)} \quad \textit{for } i = 1, \ldots, j$$

which form, therefore, a basis of subspace $H_{*p}(\mathcal{M}_p)$ of $\mathcal{N}_{H(p)}$.

We will not hesitate to use this characterization of the tangent map (i.e., use coordinate maps of this kind) when that is convenient.

Members of $\mathcal{F}^{\infty}_{H(p)}(\mathcal{N})$ may be constant on the part of $H(\mathcal{M})$ in their domains. In the situation of Theorem 17.1, where H has constant rank in the vicinity of p, when $k > j$ the coordinate gridcurves of that theorem, $G^{e_i}_{y,H(p)}$ for $i = j+1, \ldots, k$, are $k - j$ examples.

In the constant-rank case, if f is *any* member of $\mathcal{F}^{\infty}_{H(p)}(\mathcal{N})$ which is constant on the part of $H(\mathcal{M})$ in its domain then

$$df_{H(p)} = v_{j+1} dy^{j+1}_{H(p)} + \cdots + v_k dy^k_{H(p)}$$

since if $df_{H(p)}$ had any nonzero component involving $dy^1_{H(p)}, \ldots, dy^j_{H(p)}$, say for specificity $dy^1_{H(p)}$, then $f \circ G^{e_1}_{y,H(p)}$ would have nonzero derivative at 0. But $f \circ G^{e_1}_{y,H(p)} = f \circ H \circ G^{e_1}_{x,p}$, contradicting constancy on the image of H.

The following corollary is an easy consequence, characterizing the members of the subspace $H_{*p}(\mathcal{M}_p)$ as all and only those members of $\mathcal{N}_{H(p)}$ which act as the zero tangent vector when applied to smooth functions that are constant on the image of H.

17.2. ***Corollary.*** *If H has constant rank j in a neighborhood of p let*

$$S = \{\, f \in \mathcal{F}^{\infty}_{H(p)}(\mathcal{N}) \mid f(z) = 0 \text{ for every } z \text{ in both the domain of } f \text{ and } H(\mathcal{M}) \,\}.$$

$$\text{So } H_{*p}(\mathcal{M}_p) = \{ X \in \mathcal{N}_{H(p)} \mid Xf = 0 \quad \forall f \in S \,\}.$$

Many texts denote H_{*p} by dH_p and call the tangent map the differential, but we don't because we have already used that vocabulary for real valued functions f on $\mathcal{M}$, and f_{*p} and df_p are slightly different for these functions.

As we have defined it df_p is given for function f from $\mathcal{M}$ to $\mathbb{R}^m$ where $m = 1$.

If m is any positive integer we can expand that definition to $\boldsymbol{df_p} \colon \mathcal{M}_p \to \mathbb{R}^m$ given by $df_p([x,v]_p) = (f \circ x^{-1})'(p_x)\, v$. This is a linear map from $\mathcal{M}_p$ to $\mathbb{R}^m$.

df_p and f_{*p} are obviously related, and the difference corresponds to how seriously you intend to take the identity map on $\mathbb{R}^m$, whether $\mathbb{R}^m$ is to be taken on its own or as a range *manifold.*

Thus $f_{*p}([x,v]_p) = [id, df_p([x,v]_p)]_{f(p)}$ for identity map id on $\mathbb{R}^m$.

So we have found a way to transform members of $\mathcal{M}_p$ into members of $\mathcal{N}_{H(p)}$ for any $p \in \mathcal{M}$. Only points of $\mathcal{N}$ in the range of H are targeted by any of these maps.

Remember, H need not be one-to-one. If $H(p) = H(q)$ for $p \neq q$ then there will be *two* maps "pushing forward" tangent vectors into $\mathcal{N}_{H(p)} = \mathcal{N}_{H(q)}$.

If $K \colon \mathcal{N} \to \mathcal{W}$ is also a smooth map between manifolds and $p \in \mathcal{M}$ then $K \circ H(p) \in \mathcal{W}$. So we can define the tangent map of $K \circ H$ at p.

It is straightforward to show that

$$(K \circ H)_{*p} = K_{*H(p)} \circ H_{*p}.$$

This map pushes tangent vectors from $\mathcal{M}_p$ into tangent vectors in $\mathcal{W}_{K \circ H(p)}$, and we can "push forward" tangent vectors as far as we want this way.

We have defined the tangent map applied to $[h]_p$ as $[H \circ h]_{H(p)}$ and this has a serious advantage: no coordinates are directly involved in this application.

We have defined the tangent map on a coordinate representation of a tangent vector, $[x, v]_p$, as $[y, (y \circ H \circ x^{-1})'(p_x)\, v]_{H(p)}$ and this has the advantage that it tells us exactly what we need to calculate to implement the pushforward.

If you take a random walk through the literature you are likely to find formulas that don't look like this at all but which purport to define the tangent map or differential of H as well. It is a notation problem, founded in historical choices that made sense at the time, rendered ineradicable by long-use and ubiquity.

The story behind this version goes as follows.

A tangent vector at p is a derivation on $\mathcal{F}_p^\infty(\mathcal{M})$. Usually the "partial derivative" notation is employed in older texts. Tangent vectors there have the form $v^k \frac{\partial}{\partial x^k}\Big|_p$.

Derivations are determined by how they act on real-valued functions. So to push a tangent vector at p to a tangent vector at $H(p)$ we need to prescribe what the result does to a generic member of $\mathcal{F}_{H(p)}^\infty(\mathcal{N})$.

Select $f \in \mathcal{F}_{H(p)}^\infty(\mathcal{N})$. You may see a calculation similar to

$$\begin{aligned} H_{*p}\left(v^k \frac{\partial}{\partial x^k}\Big|_p \right)(f) &= v^k\, \frac{\partial}{\partial x^k}\Big|_p (f \circ H) = v^k\, \frac{\partial f}{\partial y^i}(H(p))\, \frac{\partial (y^i \circ H)}{\partial x^k}(p) \\ &= v^k\, \frac{\partial (y^i \circ H)}{\partial x^k}(p)\, \frac{\partial}{\partial y^i}\Big|_{H(p)}(f) \end{aligned}$$

followed by the conclusion that

$$H_{*p}\left(v^k\, \frac{\partial}{\partial x^k}\Big|_p \right) = \left(v^k\, \frac{\partial (y^i \circ H)}{\partial x^k}(p) \right) \frac{\partial}{\partial y^i}\Big|_{H(p)}.$$

Using the modern partial derivative notation allows us to make sense of this.

$$\begin{aligned} H_{*p}\left(v^k \frac{\partial}{\partial x^k}\Big|_p \right)(f) &= v^k D_k(f \circ H \circ x^{-1})(p) = v^k D_k(f \circ y^{-1} \circ y \circ H \circ x^{-1})(p) \\ &= v^k D_i(f \circ y^{-1})(y \circ H(p))\, D_k(y^i \circ H \circ x^{-1})(p) \\ &= \big(v^k D_k(y^i \circ H \circ x^{-1})(p) \big)\, D_i(f \circ y^{-1})(y(H(p))). \end{aligned}$$

Though identical in meaning and (after a bit of unpacking) explicit, I personally find this laborious notation and expanded calculation to lack several qualities I look for in good notation. It was even worse before the advent of the Einstein summation convention. And the notation fails to compensate for its visual clutter by guiding me to a better understanding of the nature of a tangent vector.

I recommend sticking to the earlier approach when possible.

18. Pullback on Cotangent Spaces

In this section we use the same setup as in the last: $\mathcal{M}$ and $\mathcal{N}$ are manifolds and $H\colon \mathcal{M} \to \mathcal{N}$ is a smooth map. x and y are coordinate maps around $p \in \mathcal{M}$ and $H(p) = H(p) \in \mathcal{N}$, respectively.

Suppose $g \in \mathcal{F}^{\infty}_{H(p)}$, the real-valued functions defined and smooth on some neighborhood of $H(p)$ in $\mathcal{N}$. Then $g \circ H \in \mathcal{F}^{\infty}_{p}$.

And if $(g \circ y^{-1})'(H(p)_y) = (f \circ y^{-1})'(H(p)_y)$ for some $f \in \mathcal{F}^{\infty}_{H(p)}$ then

$$\begin{aligned}(g \circ H \circ x^{-1})'(p_x) =&(g \circ y^{-1} \circ y \circ H \circ x^{-1})'(p_x)\\ =&(g \circ y^{-1})'(H(p)_y)\,(y \circ H \circ x^{-1})'(p_x)\\ =&(f \circ y^{-1})'(H(p)_y)\,(y \circ H \circ x^{-1})'(p_x) = (f \circ H \circ x^{-1})'(p_x).\end{aligned}$$

That means $d(g \circ H)_p = d(f \circ H)_p$.

So H can be used to identify in each of its domain cotangent spaces $\mathcal{M}^*_p$ a covector related (through H) to each covector in its range cotangent space $\mathcal{N}^*_{H(p)}$.

We define for each p and smooth function $H\colon \mathcal{M} \to \mathcal{N}$

$$\boldsymbol{H^*_p\colon \mathcal{N}^*_{H(p)} \to \mathcal{M}^*_p} \quad \text{by} \quad H^*_p(\, dg_{_{H(p)}}\,) = d(g \circ H)_p.$$

H^*_p is called the **pullback of H at p**. One **pulls back** covectors from $\mathcal{N}^*_{H(p)}$ into $\mathcal{M}^*_p$. The pullback is linear. The dimension of the range in $\mathcal{M}^*_p$ is the rank of H at p.

As with pushforward, the pullback process can be extended through compositions.

If $K\colon \mathcal{N} \to \mathcal{W}$ is smooth and $p \in \mathcal{M}$ then $K \circ H(p) \in \mathcal{W}$.

So at each p the pullback of $K \circ H$ at p

$$K \circ H(p)^*\colon \mathcal{W}^*_{K\circ H(p)} \to \mathcal{M}^*_p \quad \text{is} \quad (K \circ H)^*_p(\, dg_{_{K\circ H(p)}}\,) = d(g \circ K \circ H)_p.$$

A quick examination shows (note reversal of order) that

$$(K \circ H)^*_p = H^*_p \circ K^*_{H(p)}.$$

Select cotangent vector $\omega_{H(p)} = df_{_{H(p)}} = [y, \sigma]_{H(p)} \in \mathcal{N}^*_{H(p)}$.

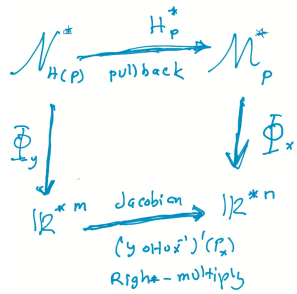
So $\sigma = (f \circ y^{-1})'(H(p)_y)$.

As we did with the tangent map, we want to calculate the value $H^*_p([y,\sigma]_{H(p)}) \in \mathcal{M}^*_p$ of the pullback map applied to this cotangent vector.

It will be of the form $[x, \tau]_p$, where $\tau = (f \circ H \circ x^{-1})'(p)$.

$$\begin{aligned}H^*_p([y,\sigma]_{H(p)}) &= [x, (f \circ y^{-1} \circ y \circ H \circ x^{-1})'(p_x)]_p\\ &= [x, \sigma\,(y \circ H \circ x^{-1})'(p_x)]_p = \left[x, \sigma\,\frac{d\,y \circ H}{dx}\right]_p = [x,\tau]_p.\end{aligned}$$

In words, you right-multiply $\boldsymbol{\sigma}$ by the Jacobian of $\boldsymbol{H}$ to get $\boldsymbol{\tau}$.

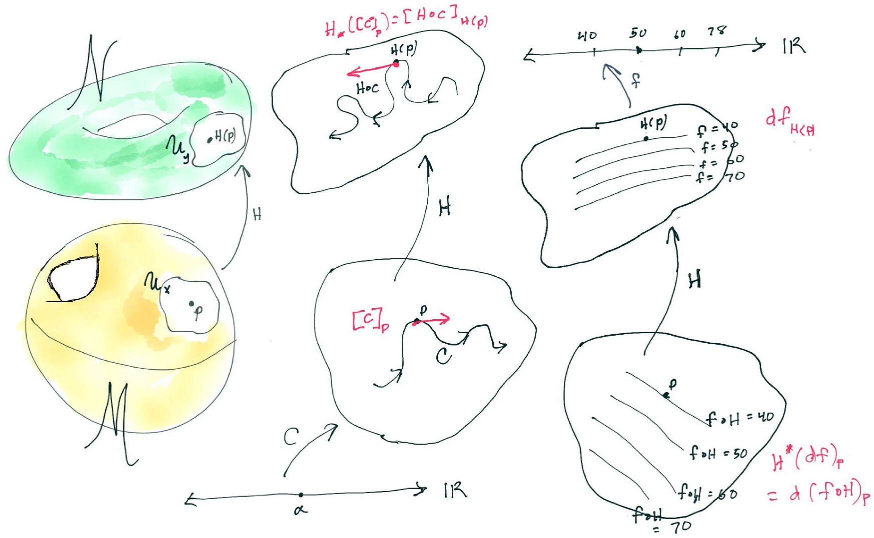


Members of tangent and cotangent spaces are determined by how they act on each other. Suppose $[c]_p = X_p \in \mathcal{M}_p$ and $df_{_{H(p)}} = \omega_{_{H(p)}} \in \mathcal{N}^*_{H(p)}$.

$$\begin{aligned} df_{_{H(p)}}\, H_{*p}([c]_p) &= df_{_{H(p)}}\, [H\circ c]_{H(p)} \\ &= (f\circ H\circ c)'(0) = d(f\circ H)_p\, [c]_p = H^*_p(df_{_{H(p)}})\, [c]_p. \end{aligned}$$

Pulling back $df_{H(p)}$ to act on $[c]_p$ gives the same result as using $df_{_{H(p)}}$ directly to act on the image of $[c]_p$ pushed into $\mathcal{N}_{H(p)}$ by the tangent map.

$$\boldsymbol{\omega_{_{H(p)}}\, H_{*p}(X_p) = H^*_p(\omega_{_{H(p)}})\, X_p.}$$



As with tangent maps some sources use an older, more cumbersome, notation for pullbacks. You will see a formula similar to

$$H^*_p(\ \sigma_i\, dy^i_{_{H(p)}}) = \left(\sigma_i\ \frac{\partial(y^i\circ H)}{\partial x^k}(p)\right) dx^k_p.$$

When $\sigma_i\, dy^i_{_{H(p)}} = df_{_{H(p)}}$ for smooth real-valued function f defined around $H(p)$ we have $\sigma_i = \frac{\partial f}{\partial y^i}(H(p)) = D_i(f\circ y^{-1})(y(H(p)))$ for each i and the chain rule gives

$$H^*_p(\ df_{_{H(p)}}) = \frac{\partial(f\circ H)}{\partial x^k}(p)\, dx^k_p = d(f\circ H)_p.$$

This can all be obtained easily from our definitions, of course, but in older texts it may be obtained by reasoning that the pullback of a cotangent vector is determined by what it does to curves through p. Let c be a smooth curve with $c(0) = p$ and apply $\sigma_i \, dy^i_{H(p)}$ to the smooth curve $H \circ c$ through $H(p)$. The chain rule yields the same result as the right-hand side applied to c, thereby justifying the formula or providing an interpretation of the definition, depending on where this formula first appears in the pullback discussion.

19. Pushforward and Pullback on Tangent and Cotangent Manifolds

We now will extend this concept to apply to tangent and cotangent *manifolds*, not just individual tangent and cotangent spaces.

The map H_{*p} has been defined on $\mathcal{M}_p$ for each p. So we can define a map

$$\boldsymbol{H_*\colon \mathcal{T}(\mathcal{M}) \to \mathcal{T}(\mathcal{N})} \quad \text{given by} \quad H_*(X_p) = H_{*p}(X_p)$$

and this is a smooth map between the two tangent manifolds.

If H is a diffeomorphism so is H_*.

Differentiable curves in any manifold are important entities, and they are important in tangent manifolds too.

Suppose $c\colon (a, b) \to \mathcal{M}$ is any smooth curve in $\mathcal{M}$.

Then the function $c_*\colon \mathcal{T}(\,(a, b)\,) \to \mathcal{T}(\mathcal{M})$ is defined by $c_*(\,[h]_\alpha\,) = [c \circ h]_{\alpha_c}$ for each $\alpha \in (a, b)$.

The nicest possible non-constant smooth curve in the manifold (a, b) corresponds to the identity map id on (a, b), and $[id, 1]_\alpha = \frac{d}{dt}\big|_\alpha$ is tangent to the identity map at every $\alpha \in (a, b)$.

The curve $b\colon (a, b) \to \mathcal{T}(\,(a, b)\,)$ given by $b(\alpha) = [id, 1]_\alpha = \frac{d}{dt}\big|_\alpha$ would have to qualify for consideration as the nicest curve in $\mathcal{T}(\,(a, b)\,)$.

For each $\alpha \in (a, b)$ let k^α denote the time-shifted curve given for each t in $(a, b) \cap (a - \alpha,\, b - \alpha)$ by $k^\alpha(t) = c(\alpha + t)$. We use this translate to produce a curve in (a, b) with the same derivative at $t = 0$ as that possessed by c at $t = \alpha$.

Composing c_* with b gives a smooth curve in $\mathcal{T}(\mathcal{M})$:

$$c_* \circ b(\alpha) = c_*\left(\frac{d}{dt}\bigg|_\alpha \right) = [k^\alpha]_{c(\alpha)} = [x, (x \circ c)'(\alpha)]_{c(\alpha)}.$$

Tangent vector $c_* \circ b(\alpha)$ is tangent to c at every α. If $\mathcal{M}$ happens to be $\mathbb{R}^n$ this can be visualized as an arrow attached to the curve c "pointing the way" along the curve, and whose length is the speed of the motion.

There are issues trying to use the point-by-point pullback maps to create a map $H^*\colon \mathcal{T}^*(\mathcal{N}) \to \mathcal{T}^*(\mathcal{M})$ between the two cotangent manifolds, as we just did for tangent manifolds.

First, the map cannot be defined at all off $H(\mathcal{M})$ so if H is not onto $\mathcal{N}$ we have one problem.

Second, if H is not one-to-one the map is not well-defined: when $H(p) = H(q)$ for $p \neq q$ we would have $H^*(\omega_{H(p)})$ multiply defined as a member of both $\mathcal{M}_p^*$ and $\mathcal{M}_q^*$.

But if H is a is one-to-one and onto all is well, and

$$\boldsymbol{H^* \colon \mathfrak{T}^*(\mathcal{N}) \to \mathfrak{T}^*(\mathcal{M})} \quad \text{can be given by} \quad H^*\left(\omega_{H(p)}\right) = H_p^*\left(\omega_{H(p)}\right).$$

As with H_*, if H is a diffeomorphism so is H^*.

20. Vector Fields and 1-Forms

A function $X \colon \mathcal{M} \to \mathfrak{T}(\mathcal{M})$ for which $X_p \in \mathcal{M}_p$ for all $p \in \mathcal{M}$ is called a **section** of the tangent bundle, and a **vector field** on $\mathcal{M}$.

Be warned of a notation expansion. Previously, X_p stood for a generic member of $\mathcal{M}_p$, and the p subscript served as a reminder of "where it was at." Now X is a function whose value at each p happens to lie in $\mathcal{M}_p$ and this function value is denoted $X_p = X(p)$.

For each coordinate map $x \colon U_x \to R_x$ any vector field can be represented as

$$X_q = v^i(q) \left.\frac{\partial}{\partial x^i}\right|_q \quad \text{for } q \in U_x$$

and the vector field is called smooth if each coefficient function $v^i \circ x^{-1}$ is smooth for all coordinate maps in a generating atlas.

The set of smooth vector fields on $\mathcal{M}$ will be denoted $\boldsymbol{\mathfrak{T}^1(\mathcal{M})}$.

In coordinates x the formula shown above on the right actually defines a member of $\mathfrak{T}^1(U_x)$. Every member of $\mathfrak{T}^1(\mathcal{M})$ can be restricted to U_x to produce a member of $\mathfrak{T}^1(U_x)$. However it is possible that $\mathfrak{T}^1(U_x)$ may contain vector fields that are *not* restrictions[16] from vector fields on $\mathcal{M}$.

For specific q and another coordinate map $y \colon U_y \to R_y$ for which $q \in U_y \cap U_x$ the coefficients $w^i(q)$ on $\left.\frac{\partial}{\partial y^i}\right|_q$ in a representation of $X(q)$ in y coordinates is specified by the change of coordinates formula:

$$w^i(q) = \frac{\partial y^i}{\partial x^j}(q)\, v^j(q).$$

Similarly, a function $\theta \colon \mathcal{M} \to \mathfrak{T}^*(\mathcal{M})$ for which θ_p, defined as $\theta(p)$, is in $\mathcal{M}_p^*$ for all $p \in \mathcal{M}$ is called a **section** of the cotangent bundle, and a **covector field** on $\mathcal{M}$.

For each coordinate map $x \colon U_x \to R_x$ the restriction of any covector field on $\mathcal{M}$ can be represented as a covector field on U_x by

$$\theta_q = \sigma_i(q)\, dx_q^i \quad \text{for } q \in U_x$$

and as in the case of vector fields, a covector field is called smooth if each coefficient function $\sigma_i \circ x^{-1}$ is smooth for each x.

[16] Certain U_x have vector fields that would, of necessity, be discontinuous if extended to $\mathcal{M}$.

With $q \in U_y \cap U_x$ as above we have $\theta_q = \tau_i(q)\, dy^i_q$ where

$$\tau_j(q) = \sigma_i(q)\, \frac{\partial x^i}{\partial y^j}(q).$$

A smooth covector field is called a **1-form on $\mathcal{M}$** and the set of all such will be denoted $\boldsymbol{\mathcal{T}_1(\mathcal{M})}$.

Both $\mathcal{T}^1(\mathcal{M})$ and $\mathcal{T}_1(\mathcal{M})$ with the natural pointwise operations are real (infinite dimensional) vector spaces, but they are also $\mathcal{F}^\infty(\mathcal{M})$-modules, and can be made to act on each other as members of dual vector spaces at each point in the manifold, and the result of such an action is a member of $\mathcal{F}^\infty(\mathcal{M})$.

In this context, members of $\mathcal{F}^\infty(\mathcal{M})$ are sometimes called **number fields on $\mathcal{M}$**.

The differential linear operator $d\colon \mathcal{F}^1_p \to \mathcal{M}^*_p$ defined at each p induces a similar map, also called the **differential**, defined on the number fields $\mathcal{F}^\infty(\mathcal{M})$, given by

$$\boldsymbol{d\colon \mathcal{F}^\infty(\mathcal{M}) \to \mathcal{T}_1(\mathcal{M})} \;\text{ given by }\; df(p) = df_p.$$

The differential obeys the rule:

$$\boldsymbol{d(fg) = f\, dg + g\, df} \quad \text{for all } f, g \in \mathcal{F}^\infty(\mathcal{M}).$$

It is, of course, *not true in general* that every member of $\mathcal{T}_1(\mathcal{M})$ can be represented as df for some $f \in \mathcal{F}^\infty(\mathcal{M})$, and "exact" conditions under which a given covector field has such a representation is an interesting question.

We can now extend the pullback and tangent map to act on number fields, vector fields and 1-forms ***in some cases.***

Suppose $H\colon \mathcal{M} \to \mathcal{N}$ is a smooth map between manifolds.

We define $\boldsymbol{H^*\colon \mathcal{F}^\infty(\mathcal{N}) \to \mathcal{F}^\infty(\mathcal{M})}$ by $H^*(f) = f \circ H$.

This is not only a homomorphism from one real vector space to another, it is an *algebra* homomorphism between these two commutative unitary algebras: that is, $H^*(fg) = H^*(f)H^*(g)$ and $H^*(1_\mathcal{N}) = 1_\mathcal{M}$, where 1_B denotes a function that is constantly 1 on set B.

Note that any two members of $\mathcal{F}^\infty(\mathcal{N})$ which agree on the image $H(\mathcal{M})$ of H will be sent to the same member of $\mathcal{F}^\infty(\mathcal{M})$. So if $H(\mathcal{M}) \neq \mathcal{N}$ then H^* will likely[17] have nontrivial kernel. But if H is a diffeomorphism H^* is an isomorphism.

For 1-form $\omega \in \mathcal{T}_1(\mathcal{N})$ define member $H^*(\omega)$ of $\mathcal{T}_1(\mathcal{M})$ to be the 1-form given on $\mathcal{M}$ by $H^*(\omega)_p = H^*_p(\omega_{H(p)})$ for each $p \in \mathcal{M}$.

Note that only the part of ω defined on the range $H(\mathcal{M}) \subset \mathcal{N}$ of H is relevant to this definition. If θ and ω agree on $H(\mathcal{M})$ then $H^*(\omega) = H^*(\theta)$.

This is **the pullback of H**,

$$\boldsymbol{H^*\colon \mathcal{T}_1(\mathcal{N}) \to \mathcal{T}_1(\mathcal{M})}.$$

[17] It is *possible* for H^* to have trivial kernel on $\mathcal{F}^\infty(\mathcal{N})$ even if H is not onto. Let $\mathcal{M}$ be the unit circle with the point $(1,0)$ removed and let $\mathcal{N}$ be the unit circle, both with subspace topology from the plane. Let H be the inclusion map of $\mathcal{M}$ into $\mathcal{N}$. Then H^* is one-to-one on $\mathcal{F}^\infty(\mathcal{N})$. However it is not onto $\mathcal{F}^\infty(\mathcal{M})$, which contains functions with different limits as you head toward $(1,0)$ from above or from below along the circle.

So if $\omega \in \mathcal{T}_1(\mathcal{N})$ and $\omega_{_{H(p)}} = dg_{_{H(p)}}$ then

$$\boldsymbol{H^*(\omega)_p = H_p^*(\omega_{_{H(p)}}) = d(g \circ H)_p.}$$

We summarize the pertinent facts now.

When y and x are coordinate maps around $H(p)$ and p, respectively, and if $\omega_{H(p)} = [y, \sigma]_{H(p)}$ where $\sigma = (g \circ y^{-1})'(H(p)_y)$ for some real-valued function g, then

$$\begin{aligned} \boldsymbol{H^*(\omega)_p} &\boldsymbol{= H_p^*([y, \sigma]_{H(p)}) = [x, \sigma\,(y \circ H \circ x^{-1})'(p_x)]_p} \\ &\boldsymbol{= [x, (g \circ y^{-1})'(H(p)_y)\,(y \circ H \circ x^{-1})'(p_x)]_p} \\ &\boldsymbol{= [x, (g \circ H \circ x^{-1})'(p_x)]_p = d(g \circ H)_p.} \end{aligned}$$

To apply H^* directly to $[y, \sigma]_{H(p)}$ you move from $\mathcal{N}$ to $\mathcal{M}$ and right-multiply σ by the Jacobian of H. Other equalities are just the chain rule.

This is ***a second usage of the symbol*** H^*. Formerly it was a map from one cotangent manifold to another. Here it is a map from covector fields to covector fields. Context indicates which is intended!

H^* cannot be an $\mathcal{F}^\infty(\mathcal{N})$-module homomorphism for a number of reasons, such as the domain mismatch between members of $\mathcal{F}^\infty(\mathcal{N})$ and members of $\mathcal{F}^\infty(\mathcal{M})$. However if $f \in \mathcal{F}^\infty(\mathcal{N})$ then $f \circ H \in \mathcal{F}^\infty(\mathcal{M})$ and the point-by-point definition of H^* gives us the relation

$$H^*(f\,\omega) = H^*(f)\,H^*(\omega) \quad \text{for } f \in \mathcal{F}^\infty(\mathcal{N}) \text{ and } \omega \in \mathcal{T}_1(\mathcal{N}).$$

We now suppose* H *is one-to-one and onto.

For vector field $X \in \mathcal{T}^1(\mathcal{M})$ define member $H_*(X)$ of $\mathcal{T}^1(\mathcal{N})$ to be the vector field on $\mathcal{N}$ given by $H_*(X)_{H(p)} = H_{*p}(X_p)$ for each $p \in \mathcal{M}$.

Had H not been onto this would only serve to define the vector field on the range of H, not all of $\mathcal{N}$. Had H not been one-to-one the function would fail to be well-defined.

The map H_* is the **pushforward or tangent map of H**,

$$\boldsymbol{H_* \colon \mathcal{T}^1(\mathcal{M}) \to \mathcal{T}^1(\mathcal{N}).}$$

So if $X \in \mathcal{T}^1(\mathcal{M})$ and curve $c \in X_p$ then

$$\boldsymbol{H_*(X)_{H(p)} = H_{*p}(X_p) = [H \circ c]_{H(p)}.}$$

Here again we have ***a second usage of the symbol*** H_*. Formerly it was a map from one tangent manifold to another. Here we take vector fields to vector fields. Context will make clear which we intend.

If we want to calculate this in coordinates select coordinate maps y and x around $H(p)$ and p, respectively. If $X_p = [x, v]_p$ where $v = v(p) = (x \circ c)'(0)$ for some curve $c \in X_p$ then

$$\begin{aligned}\boldsymbol{H_*(X)_{H(p)}} &= \boldsymbol{H_{*p}([x,v]_p) = [y, (y\circ H\circ x^{-1})'(p_x)\, v]_{H(p)}}\\ &= \boldsymbol{[y, (y\circ H\circ x^{-1})'(p_x)\,(x\circ c)'(0)]_{H(p)}}\\ &= \boldsymbol{[y, (y\circ H\circ c)'(0)]_{H(p)} = [H\circ c]_{H(p)}}.\end{aligned}$$

To apply H_* directly to $[x,v]_p$ you move from $\mathcal{M}$ to $\mathcal{N}$ and left-multiply v by the Jacobian of H. The rest is the chain rule.

If $f \in \mathcal{F}^\infty(\mathcal{M})$ and in case H is a diffeomorphism then $f\circ H^{-1} \in \mathcal{F}^\infty(\mathcal{N})$.

So we can define a map $\boldsymbol{H_*\colon \mathcal{F}^\infty(\mathcal{M}) \to \mathcal{F}^\infty(\mathcal{N})}$ by $H_*(f) = f\circ H^{-1}$.

This map H_* is an algebra isomorphism of $\mathcal{F}^\infty(\mathcal{M})$ onto $\mathcal{F}^\infty(\mathcal{N})$.

As before, the point-by-point definition of H_* gives us the relation

$$H_*(f\,X) = H_*(f)\,H_*(X) \quad \text{for } f\in\mathcal{F}^\infty(\mathcal{M}) \text{ and } X\in\mathcal{T}^1(\mathcal{M}).$$

If H is a diffeomorphism H^* and H_* are vector space isomorphisms. H_* and $H_*^{-1} = (H_*)^{-1}$ serve to identify $\mathcal{T}^1(\mathcal{M})$ with $\mathcal{T}^1(\mathcal{N})$, and the two pullbacks H^* and $(H^{-1})^* = (H^*)^{-1}$ identify $\mathcal{T}_1(\mathcal{N})$ with $\mathcal{T}_1(\mathcal{M})$.

Because of the way we defined tangent vectors and differentials the tangent maps and pullbacks of coordinate maps are trivial to apply.

When $x\colon U_x \to R_x$ is a coordinate map $\mathcal{T}(U_x)$ may be regarded as an open submanifold of $\mathcal{T}(\mathcal{M})$ and $\mathcal{T}^*(U_x)$ as an open submanifold of $\mathcal{T}^*(\mathcal{M})$.

Any diffeomorphism whose domain is an open submanifold of $\mathcal{M}$ is called a **neighborhood diffeomorphism** on $\mathcal{M}$. (Note the distinction between a neighborhood diffeomorphism and a local diffeomorphism. The latter are defined on all of $\mathcal{M}$ and may not be one-to-one.) Since x is a diffeomorphism from submanifold U_x of $\mathcal{M}$ onto submanifold R_x of $\mathbb{R}^n$ it is an example.

The tangent maps $x_*\colon \mathcal{T}^1(U_x) \to \mathcal{T}^1(R_x)$ and pullbacks $x^*\colon \mathcal{T}_1(R_x) \to \mathcal{T}_1(U_x)$ are vector space isomorphisms and *almost* $\mathcal{F}^\infty$-module isomorphisms, except for the domain mismatch issue mentioned above.

Suppose $X\in\mathcal{T}^1(U_x)$ and $\theta\in\mathcal{T}_1(R_x)$.

There is a curve c in $\mathcal{M}$ with $c(0)=p$ and function $g\colon R_x\to\mathbb{R}$ for which

$$[c]_p = X_p \text{ and } \; dg_{p_x} = \theta_{p_x}.$$

Define column vector $v = (x\circ c)'(0)$ and row vector $\tau = g'(p_x)$.

Thus $X_p = [c]_p = [x,v]_p$ and $\theta_{p_x} = \; dg_{p_x} = [id,\tau]_{p_x}$.

Applying the tangent map x_* to X at p we have

$$x_*(X)_{p_x} = x_{*p}([x,v]_p) = x_{*p}(X_p) = [x\circ c]_{p_x} = [id,(x\circ c)'(0)]_{p_x} = [id,v]_{p_x}.$$

For the pullback we have,

$$x^*(\theta)_p = x_p^*(\theta_{p_x}) = x_p^*([id,\tau]_{p_x}) = x_p^*(dg_{p_x}) = d(g\circ x)_p = [x,\tau]_p.$$

Moving around among these vector and covector fields involves nothing more than switching the identity map in R_x with x in U_x and the point p_x with p when using the "ordered pair" version of the vector and covector field values.

Theorems about vector or covector fields, such as theorems about solutions of differential equations determined by a vector field in $\mathbb{R}^n$, have local variants on manifolds, transported there by x_* and x^* and their inverses.

For neighborhood diffeomorphism x from manifold $\mathcal{M}$ onto an open subset of $\mathbb{R}^n$ and p in the domain of x

$$x_p^*([id,\tau]_{p_x}) = [x,\tau]_p \quad \text{and} \quad x_{*p}([x,v]_p) = [id,v]_{p_x}.$$

Applying $(x^{-1})^* = (x^*)^{-1}$ and $(x^{-1})_* = (x_*)^{-1}$ can be useful too, and the boxed equations help with that as well.

21. Special Cases of Tangent Map and Pullback

We will discuss these constructions involving smooth functions

$$f\colon \mathcal{M} \to \mathbb{R} \quad \text{and smooth curve} \quad c\colon \mathbb{R} \to \mathcal{M}$$

and a point $p \in \mathcal{M}$ and $r \in \mathbb{R}$ with $f(p) = r$ and $c(0) = p$.

$\mathbb{R}_0$ is the equivalence classes of real-valued functions on $\mathbb{R}$ with common derivative at time 0, which is when they pass through 0. If $\frac{d}{dt}\big|_\alpha$ is the tangent vector of the identity map id on $\mathbb{R}$ then

$$[h]_0 = h'(0)\, \frac{d}{dt}\bigg|_0 = [id, h'(0)]_0.$$

The tangent map $c_{*0}\colon \mathbb{R}_0 \to \mathcal{M}_p$ takes $[h]_0$ to $[c \circ h]_p$. The Jacobian of c with respect to coordinates x around p and id around 0 is

$$\begin{aligned} &(x \circ c \circ id^{-1})'\,(id(0)) = (x \circ c)'(0) \\ \text{so} \qquad c_{*0}([h]_0) &= c_{*0}([id, h'(0)]_0) = [x, (x \circ c)'(0) h'(0)]_p \\ &= h'(0)\,[x, (x \circ c)'(0)]_p = h'(0)[c]_p. \end{aligned}$$

These tangent vectors are all numerical multiples of the single tangent vector at p determined by c itself.

Now lets examine the pullback $c_0^*\colon \mathcal{M}_p^* \to \mathbb{R}_0^*$.

Each pullback value should be characterized as some numerical multiple of the differential dt_0 of the identity, since the range of this pullback is 1-dimensional.

If $dg_p \in \mathcal{M}_p^*$ for $g\colon \mathcal{M} \to \mathbb{R}$ then $c_0^*(dg_p) = d(g \circ c)_0$.

$$\begin{aligned} c_0^*(dg_p) &= c_0^*([x, (g \circ x^{-1})'(p_x)]_p) = [id, (g \circ c \circ id^{-1})'(id(0))]_0 \\ &= [id, (g \circ c)'(0)]_0 = (g \circ c)'(0)\,[id, 1]_0 = (g \circ c)'(0)\, dt_0. \end{aligned}$$

So in the identity coordinate map the numerical multiple is just the directional derivative of g in the direction determined by c at p.

Now let's find the tangent map and pullback for $f\colon \mathcal{M} \to \mathbb{R}$.

If $[h]_p \in \mathcal{M}_p$ and $h(0) = p$ then the tangent map $f_{*p}\colon \mathcal{M}_p \to \mathbb{R}_r$ will take $[h]_p = [x, (x \circ h)'(0)]_p$ to

$$[id, (id \circ f \circ h)'(0)]_r = [id, (f \circ h)'(0)]_r = (f \circ h)'(0) \left.\frac{d}{dt}\right|_r .$$

If f represented something like temperature on the manifold, the numerical multiple is the rate at which temperature is rising as you move through p on h. The numerical coefficient is $df_p\,[h]_p$.

This highlights the distinction between the closely related maps df_p and f_{*p}. They both are linear maps operating on tangent vectors, but their range is different. One produces a tangent vector while the other produces a number, the *coordinate* of that tangent vector in a preferred basis.

$$f\colon \mathcal{M} \to \mathbb{R} \quad \Longrightarrow \quad f_{*p}(X_p) = [id,\, df_p(X_p)]_r = df_p(X_p) \left.\frac{d}{dt}\right|_r .$$

Finally, we come to the pullback $f_p^*\colon \mathbb{R}_r^* \to \mathcal{M}_p^*$. Suppose h is a real valued function defined on an interval containing r.

$$\begin{aligned} f_p^*(dh_r) &= f_p^*([t, (h \circ t^{-1})'(r)]_r) = f_p^*([t, h'(r)]_r) = [x, (h \circ f \circ x^{-1})'(p_x)]_p \\ &= [x, h'(r)(f \circ x^{-1})'(p_x)]_p = h'(r)\,[x, (f \circ x^{-1})'(p_x)]_p = h'(r)\,df_p. \end{aligned}$$

So $f_p^*(dh_r)$ is always a numerical multiple of df_p.

As a special case, if $h = id$ we have $f_p^*(dt_r) = df_p$. This pullback takes the differential dt of the identity map in $\mathbb{R}$ to df.

22. Differential Equations on an Open Submanifold of $\mathbb{R}^n$

At this point we have all we need to solve differential equations on manifolds, as we did in $\mathbb{R}^n$ in Section 3.

We start with an open submanifold R of $\mathbb{R}^n$ and a particularly nice vector field $Z \in \mathcal{T}^1(R)$.

For this vector field Z there is a smooth vector-valued function g on R for which

$$Z_a = [id_R,\, g(a)]_a \quad \text{for every } a \in R$$

where id_R is the identity coordinate map on R.

According to the results in Section 3 for each $p \in R$ there are numbers r and K and a smooth local flow

$$S\colon [-2r, 2r] \times B_p(2K) \to B_p(4K) \subset R$$

for the DE determined by the smooth vector-valued function g.

For each $a \in B_p(2K)$ the function $c_a\colon [-2r, 2r] \to B_p(4K)$ given by $c_a(t) = S_t(a)$ is differentiable for all t in $(-2r, 2r)$ and for these t values

$$c_a(0) = a \quad \text{and} \quad \dot{c}_a(t) = g(c_a(t)).$$

S is unique in the sense that any solution to this IVP must agree with the one provided by S on the overlap of their domain time intervals.

In Section 3, the number r was chosen so that

$$c_a(t) \in B_p(2K) \quad \text{for} \quad a \in B_p(K) \quad \text{and} \quad t \in (-r, r).$$

For open ball $B_p^o(K)$ and $t \in (-2r, 2r)$ define open set $R_t = S_t(B_p^o(K))$.

So $R_0 = B_p^o(K)$ and $R_t = S_t(B_p^o(K)) \subset B_p^o(2K) \subset R$ for all $t \in (-r, r)$.

The restricted map $S_t \colon R_0 \to R_t$ is a diffeomorphisms for each $t \in (-2r, 2r)$.

And for any $s, t \in (-r, r)$ and any $a \in B_p^o(K)$ we have

$$S_0(a) = a \quad \text{and} \quad S_{s+t}(a) = S_t(S_s(a)) \quad \text{and} \quad S_s(R_t) = R_{s+t}.$$

For any $t \in (-r, r)$ and $a \in B_p^o(K)$ it follows that

$$\dot{c}_a(t) = \lim_{\varepsilon \to 0} \frac{S_{t+\varepsilon}(a) - S_t(a)}{\varepsilon} = g(S_t(a)) = g(c_a(t)) = g(\, c_{S_t(a)}(0)\,) = \dot{c}_{S_t(a)}(0).$$

Among other things, this means for a close to p

$$Z_a = [id_R,\, g(a)]_a = \left[id_R,\, \lim_{\varepsilon \to 0} \frac{S_\varepsilon(a) - a}{\varepsilon} \right]_a = [id_R,\, \dot{c}_a(0)]_a\,.$$

The following result has a certain visual appeal.

22.1. ***Theorem.*** *If $p = 0 \in R$ and $Z_0 = [id_R,\, e_1]_0$ where e_1 is the first basis vector in $\mathbb{R}^n$ then there is a coordinate map $y \colon U_y \to R_y$ around 0 so that*

$$Z_q = [y, e_1]_q \quad \forall q \in U_y.$$

In other words, $Z_q = \frac{\partial}{\partial y^1}\Big|_q$ for all q in this coordinate patch, and on the entire patch the integral curves are just the coordinate gridcurves, $G^{e_1}_{y,q}$.

Proof. Assume the conditions on Z and that for smooth vector valued g

$$Z_a = [id_R, g(a)]_a \qquad \text{for } a \in R.$$

Let $S \colon [-2r, 2r] \times B_0(2K) \to B_0(4K)$ be the local flow for Z as discussed above.

For each a the curve $c_a(t) = S_t(a)$ is a solution curve for the DE determined by g with initial condition $c_a(0) = a$, and using the other vocabulary c_a is an integral curve for the vector field Z. In fact $[c_a]_a = Z_a$ so $\dot{c}_a(0) = g(a)$ for every a.

Suppose β is the lesser of K or r and $a \in B_0^o(\beta)$.

We temporarily suspend, for typographical convenience, our insistence that members of $\mathbb{R}^n$ be represented by columns.

Define the function ϕ by

$$\phi(a) = S_{a^1}(0, a^2, \ldots, a^n) = S_{a_1}(\, a - a^1\, e_1\,).$$

To calculate $\phi(a)$ start from the projection of a onto the first-coordinate-zero-hyperplane and follow the integral curve (the solution to the IVP) for time a^1.

Note that $\phi(0) = 0$ and, in fact, $\phi(0, a_2, \ldots, a_n) = (0, a_2, \ldots, a_n)$.

For every a in the open ball $B_0^o(\beta)$

$$
\begin{aligned}
D_1\phi(a) &= \lim_{\varepsilon\to 0}\frac{\phi(a^1+\varepsilon, a^2,\dots,a^n)-\phi(a)}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{S_{a^1+\varepsilon}(0,a^2,\dots,a^n)-\phi(a)}{\varepsilon}\\
&= \lim_{\varepsilon\to 0}\frac{S_\varepsilon(\,S_{a^1}(0,a^2,\dots,a^n)\,)-\phi(a)}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{S_\varepsilon(\phi(a))-\phi(a)}{\varepsilon}\\
&= g(\phi(a)).
\end{aligned}
$$

We don't have as much information about the partial derivatives in other directions, but we *can* do calculations at $(0, a_2, \dots, a_n)$. Specifically, for all other partial derivatives

$$
\begin{aligned}
D_j\phi(0,a_2,\dots,a_n) &= \lim_{\varepsilon\to 0}\frac{\phi(0,a_2,\dots,a_j+\varepsilon,\dots,a_n)-\phi(0,a_2,\dots,a_n)}{\varepsilon}\\
&= \lim_{\varepsilon\to 0}\frac{(0,a_2,\dots,a_j+\varepsilon,\dots,a_n)-(0,a_2,\dots,a_n)}{\varepsilon}\\
&= e_j.
\end{aligned}
$$

In particular, the derivative matrix of ϕ at the origin is the $n\times n$ identity matrix. And the first column of the derivative matrix is, everywhere on the open ball, the vector given by g at that place.

That means several things. For instance there is an open set $\mathcal{A}$ with $0\in\mathcal{A}$ upon which ϕ is a diffeomorphism between $\mathcal{A}$ and $\mathcal{B}=\phi(\mathcal{A})$. Recall that $\mathcal{A}\subset B_0^o(\beta)$ by assumption, and by restricting to smaller open neighborhoods of 0 if necessary we may without loss presume that $\mathcal{B}\subset B_0^o(\beta)$ too.

Note that the intersection of $\mathcal{A}$ and $\mathcal{B}$ with the hyperplane is the same: both ϕ and ϕ^{-1} are the identity on that intersection.

You find $\phi^{-1}(a)$ as follows. a^1 is the time parameter that got you to a from the hyperplane. Thus $S_{-a^1}(a)$ is on the hyperplane. So $\phi^{-1}(a) = a^1e_1 + S_{-a^1}(a)$.

Let's examine the form of Z on points in the coordinate patch $\mathcal{B}$ for coordinate map ϕ^{-1}.

$$\text{So}\quad Z_a = [id_{\mathcal{B}}, g(a)]_a = [\phi^{-1},\, (\phi^{-1})'(a)\, g(a)]_a$$

where $id_{\mathcal{B}}$ is the identity coordinate map on $\mathcal{B}\subset R$.

$g(a)$ is, everywhere on $\mathcal{B}$, the first column of $\phi'(a)$. Therefore $(\phi^{-1})'(a)\,g(a) = e_1$ everywhere on $\mathcal{B}$. That is:

$$Z_a = [id_{\mathcal{B}}, g(a)]_a = [\phi^{-1},\, e_1]_a \quad \text{for all } a\in\mathcal{B}.$$

□

23. Differential Equations on a Manifold

Using the results from Section 22 we will now solve differential equations on more general manifolds.

We suppose given vector field $X \in \mathcal{T}^1(\mathcal{M})$ and coordinate map $x\colon U_x \to R_x$ on manifold $\mathcal{M}$ and $p \in U_x$.

Let's restrict our attention temporarily to the open submanifold U_x of $\mathcal{M}$ and let Y be the restriction of X to this submanifold.

There is a smooth vector-valued function $v\colon U_x \to \mathbb{R}^n$ so that for all $q \in U_x$ we have $X_q = Y_q = [x,\, v(q)]_q$.

id_{R_x} denotes the identity coordinate map on R_x. Let $v = g \circ x$.

For each $q \in U_x$ $\qquad x_*(Y)_q = x_{*q}(X_q) = [id_{R_x},\, v(q)]_{q_x} = [id_{R_x},\, g(q_x)]_{q_x}$.

Let $Z = x_*(Y) \in \mathcal{T}^1(R_x)$.

$Z_{q_x} = [id_{R_x},\, g(q_x)]_{q_x}$ for every $q \in U_x$ or, as is frequently more convenient,

$$Z_a = [id_{R_x},\, g(a)]_a \quad \text{for every } a \in R_x.$$

Select K so that the closed ball $B_{p_x}(4K)$ of radius $4K$ for positive K centered at p_x is contained in R_x.

According to the results in Section 22 there is an $r > 0$ and a local flow $S\colon [-2r, 2r] \times B_{p_x}(2K) \to B_{p_x}(4K)$ for the DE determined by the vector-valued function g. The curve c_a given by $c_a(t) = S_t(a)$ is a solution curve to the IVP given by g with initial condition $c_a(0) = a$.

Now let

$$T\colon [-2r, 2r] \times x^{-1}(\, B_{p_x}(2K)\,) \to x^{-1}(\, B_{p_x}(4K)\,)$$

be defined by

$$T_t = x^{-1} \circ S_t \circ x \quad \text{for} \quad t \in [-2r, 2r].$$

$x^{-1}(\, B_{p_x}(2K)\,)$ is a compact neighborhood of p and T_0 is the identity on this set. For $t \in [-r, r]$ the restriction of T_t to $x^{-1}(\, B^0_{p_x}(K)\,)$ is a diffeomorphism of this open submanifold of $\mathcal{M}$ onto its image.

For $s, t \in [-r, r]$ and any $q \in x^{-1}(\, B^o_{p_x}(K)\,)$ we have

$$T_{s+t}(q) = T_t(T_s(q)).$$

Suppose I is an open interval and $b\colon I \to \mathcal{M}$ is a smooth curve.

b is called an **integral curve of the vector field X** and X itself is said to be **everywhere tangent to b** if $X_{b(\alpha)}$ is tangent to b at α for every $\alpha \in I$.

We have found that $X_{b_q(\alpha)}$ is tangent to the curve $b_q = x^{-1} \circ c_{q_x}$ for each $q \in x^{-1}(\, B^o_{p_x}(2K)\,)$ and at each $\alpha \in (-2r, 2r)$.

We call T a **local flow for the vector field X** and X is called an **infinitesimal generator for T**.

$$X_q = \left[x, \lim_{\varepsilon \to 0} \frac{x \circ T_\varepsilon(q) - x(q)}{\varepsilon}\right]_q = [x, (x \circ b_q)'(0)]_q \quad \text{for } q \in x^{-1}(\, B^o_{p_x}(K)\,).$$

We can carry out this procedure for any point $p \in \mathcal{M}$. There is a local flow "around" p with infinitesimal generator X for every p in $\mathcal{M}$.

23.1. ***Theorem.*** *For every smooth vector field X on manifold $\mathcal{M}$ and each point $p \in \mathcal{M}$ there are open neighborhoods $\mathcal{U}$ and $\mathcal{V}$ containing p and an interval $(-2r, 2r)$ and a local flow $T\colon (-2r, 2r) \times \mathcal{U} \to \mathcal{V}$ for X.*

By this we mean that the vector field X is everywhere tangent to

$b_q\colon (-2r, 2r) \to \mathcal{V}$ defined by $b_q(t) = T_t(q)$ for every $t \in (-2r, 2r)$.

And every curve b defined on an interval to which X is everywhere tangent and for which $b(0) = p$ agrees with b_p on some neighborhood of 0.

T is smooth and each T_t is a diffeomorphism and

$$T_{t+s} = T_s \circ T_t \quad \textit{for all} \quad s, t \in (-r, r).$$

We have just used our ability to solve differential equations in $\mathbb{R}^n$ to produce integral curves to vector fields at every point on $\mathcal{M}$. We know integral curves exist.

Suppose b_p is the integral curve of Theorem 23.1 and p is on a (possibly) different integral curve $b_1\colon I_1 \to \mathcal{M}$ for vector field X. By applying a time-shift to b_1 we produce integral curve b_2 for which $b_2(0) = p$. By restricting to a smaller open interval if necessary we may presume $b_3\colon I_3 \to U_x$. Then the curve $c_3\colon I_3 \to R_x$ defined by $c_3 = x \circ b_3$ is a solution to the IVP that was used to find c_{p_x} in the previous discussion. By the uniqueness of such solutions c_{p_x} and c_3 agree on the intersection of their domain time intervals, which contains 0, and $c_{p_x}(0) = c_1(0) = p_x$. So b_2 and b_p agree, at least on the interval $I_3 \cap (-r, r)$ around 0. This is "local uniqueness."

Now suppose $f\colon I_1 \to \mathcal{M}$ and $h\colon I_2 \to \mathcal{M}$ are two integral curves for vector field X defined on open intervals and $f(t_0) = h(t_1) = p$. If $f(t_0 + t)$ ever differs from $h(t_1 + t)$ at some positive time t there would be a greatest time $t_2 > 0$ so that $f(t_0 + t)$ and $h(t_1 + t)$ agree for $t \in [0, t_2]$. But then $f(t_0 + t_2) = h(t_1 + t_2)$, and $f(t_0 + t)$ and $h(t_1 + t)$ differ for t in an arbitrarily short open interval around t_2, contradicting local uniqueness. The same contradiction obtains if $f(t_0 + t)$ and $h(t_1 + t)$ differ at some negative time t.

We conclude that f and h must agree wherever both are defined after a single time-shift to match up the times at which they are both at p.

23.2. ***Theorem.*** *For smooth vector field X and $p \in \mathcal{M}$ there is a unique integral curve $c\colon I \to \mathcal{M}$ for X for which $0 \in I$ and $c(0) = p$ with largest open*[18] *interval domain I.*
Every integral curve through p is a time-shifted restriction of this one.

Proof. Let $\mathcal{S}$ denote the set of integral curves $b\colon I_b \to \mathcal{M}$ for X where $0 \in I_b$ and $b(0) = p$. $\mathcal{S}$ is nonempty and by the previous discussion two members of $\mathcal{S}$ agree on the intersection of their domains of definition. Let

$$I = \bigcup_{b \in \mathcal{S}} I_b.$$

[18]If you extend consideration to allow one-sided derivatives at endpoints, the interval *still* must be open if it is largest and X is everywhere tangent to c, since the local existence and uniqueness theorem would allow us to extend c beyond a closed endpoint.

I is the union of open intervals all containing 0 and so is itself an open interval.

We can now define $c\colon I \to \mathcal{M}$ by $c(t) = b(t)$ whenever $t \in I_b$. □

So how big is this interval $I = (u, v)$? Must it be all of $\mathbb{R}$?

The answer to this is no. In fact easy examples (with $\mathcal{M} = \mathbb{R}$) show that a solution curve might "go to infinity" or "go to the edge" in finite time.

In fact, that characterizes such curves. Suppose $c(I)$ is contained in some compact set K. Since $\mathcal{M}$ is locally compact we may presume there are open sets U and V each with compact closure and such that

$$c(I) \subset K \subset U \subset \overline{U} \subset V \subset \overline{V}.$$

There is a member $f\colon \mathcal{M} \to [0, 1]$ of $\mathcal{F}^\infty(\mathcal{M})$ for which

$$\overline{U} \subset f^{-1}(1) \quad \text{and} \quad \mathit{support}(f) \subset V.$$

Let $Y = fX$. Then Y agrees with X on open U so any integral curve for Y that never leaves U will be an integral curve for X too. In particular our curve c is an integral curve for Y. And Y has compact support in V.

Now for each point q in $\overline{V}$ we can create a local flow T^q where

$$T^q\colon [-2r_q, 2r_q] \times A_q \to B_q$$

where $T_0^q(z) = z$ for all $z \in A_q$, q itself is in the interior A_q^o of A_q, and $A_q \subset B_q$.

The various A_q^o form an open cover of $\overline{V}$ so we can select a finite subcover $A_{q_1}^o, \dots, A_{q_k}^o$. Let r be the least of all the $r_{q_1}, \dots, r_{q_k}$.

Now define $T\colon [-2r, 2r] \times \mathcal{M} \to \mathcal{M}$ by setting, for $t \in [-2r, 2r]$,

$$T_t(q) = q \text{ when } q \notin \overline{V} \quad \text{and} \quad T_t(q) = T_t^{q_j}(q) \text{when } q \in A_{q_j}^o.$$

Every $q \in \overline{V}$ is in at least one $A_{q_j}^o$ and by the local uniqueness of integral curves we know the values used above cannot conflict if q is also in $A_{q_i}^o$ and $i \neq j$.

Since we have T defined for a fixed short time interval $[-r, r]$ on all of $\mathcal{M}$ we can extend T to all times too.

Suppose we have defined T on all of $\mathcal{M}$ on time interval $[0, kr]$ for some positive integer k, as we have at the outset for $k = 1$. For $t \in (kr, (k+1)r]$ define

$$T_t(q) = T_{t-kr}(T_{kr}(q)) \quad \forall q \in \mathcal{M}.$$

By an induction argument we have T defined for all positive T and extend to negative t by a similar method.

Getting back to our integral curve $c\colon (u, v) \to \mathcal{M}$, we have shown if the image of c is contained in any compact set then $(u, v) = \mathbb{R}$.

Conversely, suppose $(u, v) \neq \mathbb{R}$. For specificity, say $v \neq \infty$. Then $[0, v)$ must extend beyond every compact set: either c leaves each compact set permanently after some time in $[0, v)$ or it returns and leaves again infinitely often.

The corresponding fact for negative times is true if $u \neq -\infty$.

During this discussion, we have actually proved the following theorem.

23.3. ***Theorem.*** *If the smooth vector field X has compact support there is a* ***global flow*** $T\colon \mathbb{R}\times\mathcal{M}\to\mathcal{M}$ *for X. By this we mean that for each $p\in\mathcal{M}$*

(i) $c_p\colon \mathbb{R}\to\mathcal{M}$, *defined by* $c_p(t)=T_t(p)$ *is an integral curve for X.*

(ii) Every integral curve through p is a time-shifted restriction of c_p.

(iii) T is smooth and each T_t is a diffeomorphism and T_0 is the identity on $\mathcal{M}$.

(iv) $T_{t+s}=T_s\circ T_t$ *for all s,t.*

As a final result we generalize Theorem 22.1.

23.4. ***Theorem.*** *For every smooth vector field X and every point p for which $X_p\neq 0$ there is a coordinate map $y\colon U_y\to R_y$ around p so that*

$$X_q=[y,e_1]_q \quad \forall q\in U_y. \quad (e_1 \text{ is the first basis vector in } \mathbb{R}^n.)$$

In other words, $X_q=\left.\frac{\partial}{\partial y^1}\right|_q$ *for all q in this coordinate patch, and on the entire patch the integral curves are just the coordinate gridcurves,* $G^{e_1}_{y,q}$.

Proof. To begin the discussion, suppose X is a smooth vector field for which $X_p\neq 0$ and $x\colon U_x\to R_x$ is any coordinate map around p.

There is a smooth vector-valued function $v\colon U_x\to\mathbb{R}^n$ so that for all $q\in U_x$ we have $X_q=[x,v(q)]_q$.

So for each $q\in U_x$ we create $Z_{q_x}\in(R_x)_{q_x}$ for which

$$Z_{q_x}=x_*(X)_q=x_{*q}(X_q)=[id_{R_x},v(q)]_{q_x}=[id_{R_x},g(q_x)]_{q_x}$$

where $g\circ x=v$, which determines a smooth vector field Z on R_x.

$$\text{So}\quad Z_a=[id_{R_x},g(a)]_a.$$

Modify x if necessary so that $p_x=0$ and $g(0)=e_1$: replace "old x" with a "new x" given by $A(x-p_x)$ for appropriate invertible matrix A if the "old x" doesn't satisfy these properties.

Z now satisfies the conditions of Theorem 22.1 and in that theorem we find that there is a diffeomorphism $\phi\colon\mathcal{A}\to\mathcal{B}$ where both $\mathcal{A}$ and $\mathcal{B}$ are open neighborhoods of 0 and contained in R_x and for which

$$Z_a=[id_{\mathcal{B}},g(a)]_a=[\phi^{-1},\,e_1]_a \quad \text{for all } a\in\mathcal{B}.$$

Using the handy representation of pushforward by a coordinate map we have

$$\begin{aligned}(\phi^{-1})_{*a}(Z_a)&=(\phi^{-1})_{*a}([id_{\mathcal{B}},g(a)]_a)=(\phi^{-1})_{*a}([\phi^{-1},\,e_1]_a)\\&=[id_{\mathcal{A}},\,e_1]_{\phi^{-1}(a)}\in\mathcal{A}_{\phi^{-1}(a)}.\end{aligned}$$

So let's start again with vector field X but this time restricted to open neighborhood $\mathcal{C}=x^{-1}(\mathcal{B})$ of p in $\mathcal{M}$. For q in this set we can push forward X_q to Z_{q_x} using x_* and then push forward one more time to $[id_{\mathcal{A}},\,e_1]_{\phi^{-1}\circ x(q)}$ using ϕ^{-1}_*.

The calculation is, for any $q\in\mathcal{C}$,

$$\begin{aligned}(\,(\phi^{-1}\circ x)_*(X)\,)_q&=(\phi^{-1}\circ x)_{*q}(X_q)=(\phi^{-1}\circ x)_{*q}([x,v(q)]_q)\\&=(\phi^{-1})_{*q_x}\circ x_{*q}([x,v(q)]_q)=(\phi^{-1})_{*q_x}([id_{\mathcal{B}},g(q_x)]_{q_x})\\&=(\phi^{-1})_{*q_x}([\phi^{-1},\,e_1]_{q_x})=[id_{\mathcal{A}},\,e_1]_{\phi^{-1}(q_x)}.\end{aligned}$$

$y = \phi^{-1} \circ x$ confined to coordinate patch $\mathcal{C}$ containing p is a composite coordinate map which satisfies the conclusion of the theorem.

$$X_q = [x, v(q)]_q = [y, e_i]_q \quad \forall q \in \mathcal{C}.$$

□

24. Tensor Fields

We have built a vector space $\mathcal{M}_p$ and its dual $\mathcal{M}_p^*$ at each point p on $\mathcal{M}$. Therefore we can define at each point the whole vector space structure of tensors, and can specify a type of tensor and a particular tensor of that type at each point in $\mathcal{M}$. Such a specification is called a **tensor field on $\mathcal{M}$**.

Recall that if V is any vector space we define $\mathcal{T}_s^r(V)$ to be the tensors on V covariant of order s and contravariant of order r. The set of r-forms on V is indicated by $\Lambda_r(V)$, the set of alternating members of $\mathcal{T}_r^0(V)$.

If each selection of our tensor field on $\mathcal{M}$ is in $\mathcal{T}_0^0(\mathcal{M}_p) = \Lambda_0(\mathcal{M}_p) = \mathbb{R}$ we have an ordinary real valued function on $\mathcal{M}$.

When $x\colon U_x \to R_x$ is a coordinate map on $\mathcal{M}$ then a tensor field can be represented within U_x like this:

$$T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(x)\ \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s}$$

and alternating tensors can be represented as

$$\begin{aligned} &R_{i_1,\ldots,i_r}(x)\ dx^{i_1} \otimes \cdots \otimes dx^{i_r} && (R \text{ is alternating, sum here on all indices.})\\ &\qquad = R_{i_1,\ldots,i_r}(x)\ dx^{i_1} \wedge \cdots \wedge dx^{i_r}. && (\text{Sum here on increasing indices only.}) \end{aligned}$$

where the numbers $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(x)$ and $R_{i_1,\ldots,i_r}(x)$ can, in principle, vary wildly or randomly as we move from one point to another in U_x.

However wildly they might change from point to point, if y is another coordinate map and at any particular point in $U_y \cap U_x$ the number $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(x)$ will be related to the number $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(y)$ according to the standard change-of-basis pattern for tensor coefficients.

Specifically we must have, evaluated at each point in $U_y \cap U_x$,

$$T^{w_1,\ldots,w_r}_{h_1,\ldots,h_s}(y) = T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(x)\ \frac{\partial y^{w_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{w_r}}{\partial x^{i_r}}\ \frac{\partial x^{j_1}}{\partial y^{h_1}} \cdots \frac{\partial x^{j_s}}{\partial y^{h_s}}.$$

An important case is the situation with $\Lambda_n(\mathcal{M}_p)$ for n-dimensional $\mathcal{M}$. The dimension of the space of n-forms at each p is 1: they are all multiples of any nonzero n-form. In coordinate neighborhood U_x any field of n-forms λ is a real function of position times $dx^1 \wedge dx^1 \wedge \cdots \wedge dx^n$. If y is another coordinate system then the same field can be represented on $U_x \cap U_y$ as

$$\lambda(x)\, dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n = \lambda(y)\, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^n.$$

The change of basis formula here reduces to $\lambda(y) = \lambda(x) \det\left(\frac{dx}{dy}\right)$ which can be found by examining the nonzero terms in the expanded wedge product using

$$dx^i = \frac{\partial x^i}{\partial y^1}\, dy^1 + \frac{\partial x^i}{\partial y^2} dy^2 + \cdots + \frac{\partial x^i}{\partial y^n} dy^n.$$

We are not generally interested in tensor fields that change randomly as we move across $\mathcal{M}$. Requiring the coefficient assignments to have tensor character is not enough: to capture features of most physical phenomena they should change smoothly as we move from point to point.

We will define $\boldsymbol{\mathcal{T}^r_s(\mathcal{M})}$ and $\boldsymbol{\Lambda_r(\mathcal{M})}$ to consist of those tensor fields of the indicated type for which the coordinates $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}(x)$ are **infinitely differentiable functions on $\boldsymbol{U_x}$ for every coordinate map $\boldsymbol{x}$ for $\boldsymbol{\mathcal{M}}$**. It is sufficient to require this differentiability condition for a collection of coordinate maps whose domains cover $\mathcal{M}$.

All the standard tensor constructions and operations, such as tensor and wedge products and contraction, can be performed pointwise without lowering the differentiability of the coefficients. $\mathcal{T}^r_s(\mathcal{M})$ and $\Lambda_r(\mathcal{M})$ are real vector spaces, but also $\mathcal{F}^\infty(\mathcal{M})$-modules.

Note that $\mathcal{T}^0_0(\mathcal{M}) = \Lambda_0(\mathcal{M}) = \mathcal{F}^\infty(\mathcal{M})$.

Also $\mathcal{T}^1(\mathcal{M}) = \mathcal{T}^1_0(\mathcal{M})$ and $\mathcal{T}_1(\mathcal{M}) = \mathcal{T}^0_1(\mathcal{M}) = \Lambda_1(\mathcal{M})$.

For n-manifold $\mathcal{M}$ the collection of formal sums

$$\Lambda_0(\mathcal{M}) \oplus \Lambda_1(\mathcal{M}) \oplus \cdots \Lambda_n(\mathcal{M}) \oplus \{0\} \oplus \cdots$$

with wedge product constitute the **Grassmann algebra** $\mathcal{G}(\mathcal{M})$ on $\mathcal{M}$.

It too is an $\mathcal{F}^\infty(\mathcal{M})$-module. Elements of the Grassmann algebra are called **multi-forms**, though the vocabulary is usually deployed when more than one term in the formal sum is nonzero.

An important example is a choice of volume element on each $\mathcal{M}_p$, a never-zero member of $\Lambda_n(\mathcal{M})$ for n-dimensional $\mathcal{M}$.

On the domain of coordinate map x a volume element γ can be expressed as $\gamma = \sigma\, dx^1 \wedge \cdots \wedge dx^n$ where σ is never 0. This defines an orientation on the submanifold U_x of $\mathcal{M}$ as follows. For $p \in U_x$, if $v_1, \ldots, v_n$ denotes any ordered basis of $\mathcal{M}_p$ then this ordered basis is said to have the orientation determined by γ if $\gamma(p)(v_1, \ldots, v_n) > 0$. Thus every basis of every $\mathcal{M}_p$ is one of two types. Such a basis either has the orientation determined by γ or not

Another approach to orientation is given by an atlas of coordinate maps for which the determination of orientation by $dx^1 \wedge \cdots \wedge dx^n$ on U_x is consistent on the overlap of the domains of atlas members. $\mathcal{M}$ is said to be **orientable** with orientation determined by this atlas.

A third approach is to specify directly an ordered list of smooth vector fields $X_1, X_2, \ldots, X_n$ on $\mathcal{M}$ which are linearly independent at each $p \in \mathcal{M}$. These can be used to identify charts in a consistent atlas, and thereby an orientation, as above.

Not all manifolds are orientable, though it is interesting (and not too hard to see) that all *tangent* manifolds, $\mathcal{T}(\mathcal{M})$, are orientable.

Another common construction is a choice of metric tensor $G = G_{i,j}(x)\, dx^i \otimes dx^j$ in $\mathcal{T}_2^0(\mathcal{M})$. This allows or produces constructions such as raising and lowering indices and (together with an orientation) a volume element and the Hodge star operator.

Yet another example might be a process that creates a linear function $F_p\colon \mathcal{M}_p \to \mathcal{M}_p$ for each p. It is easy to show that this corresponds to a member of $\mathcal{T}_1^1(\mathcal{M}_p)$. If this process is smooth enough we will have a member of $\mathcal{T}_1^1(\mathcal{M})$.

It is important to note (we used this before and record it here again) that submanifolds of $\mathcal{M}$ can be used to define "sub-objects" in a natural way.

If p is a point in submanifold $\mathcal{N}$ of $\mathcal{M}$ then each tangent vector $X_p \in \mathcal{N}_p$ contains only differentiable parameterizations with range in $\mathcal{N}$, while a member of $\mathcal{M}_p$ may contain curves with range outside $\mathcal{N}$. However there is one and only one member $Y_p \in \mathcal{M}_p$ with $X_p \subset Y_p$, and when we identify $\mathcal{N}_p$ with a subspace of $\mathcal{M}_p$ it is this implied association that is intended.

Also, if U is an open subset of $\mathcal{M}$, every member of $\mathcal{M}_p$ corresponds to some tangent vector in U_p. $\mathcal{T}(U)$ can be regarded by this identification as an open submanifold of $\mathcal{T}(\mathcal{M})$, and both have dimension twice that of $\mathcal{M}$.

Similarly, if U is open the covector $df(p) \in U_p^*$ contains only real valued functions with domain in U, while a member of $\mathcal{M}_p^*$ contains functions with (possibly) larger domain. However there is one and only one member of $\mathcal{M}_p^*$ containing all of $df(p)$, and we identify $df(p)$ with that member. And every member of $\mathcal{M}_p^*$ contains a unique member of U_p^* as a subset.

By this means $\mathcal{T}^*(U)$ is regarded as an open submanifold of $\mathcal{T}^*(\mathcal{M})$.

Finally, for each r and s the $\mathcal{F}^\infty(U)$-module of tensor fields $\mathcal{T}_s^r(U)$ is not contained in the $\mathcal{F}^\infty(\mathcal{M})$-module $\mathcal{T}_s^r(\mathcal{M})$. However the *restriction* of every member of $\mathcal{T}_s^r(\mathcal{M})$ to U is a member of $\mathcal{T}_s^r(U)$.

Note also, if U is a proper open subset of $\mathcal{M}$ there may be members of $\mathcal{T}_s^r(U)$ that *are not* restrictions of members of $\mathcal{T}_s^r(\mathcal{M})$. In particular, there may be members of $\mathcal{F}^\infty(U)$ that cannot be extended to a member of $\mathcal{F}^\infty(\mathcal{M})$.

25. Derivations and $\mathcal{F}^\infty(\mathcal{M})$-Module Homomorphisms

Going on to a different matter, we make another examination of derivations. A **derivation** on the number fields $\Lambda_0(\mathcal{M}) = \mathcal{F}^\infty(\mathcal{M})$ is a function $D\colon \Lambda_0(\mathcal{M}) \to \Lambda_0(\mathcal{M})$ for which for any constant function c and functions f and g in $\Lambda_0(\mathcal{M})$ we have $D(cf + g) = cD(f) + D(g)$ and $D(fg) = fD(g) + gD(f)$.

We discussed point derivations on $\mathcal{F}_p^\infty(\mathcal{M})$, the real functions smooth on some neighborhood of p, in section 11, and concluded that action by tangent vectors exhausts all possible point derivations. In fact, these point derivations are in many treatments taken *to be* the tangent vectors at p.

Evaluation of the output of a derivation at $p \in \mathcal{M}$ *looks* like it will produce a point derivation on each $\mathcal{F}_p^\infty(\mathcal{M})$. Specifically, if $g \in \Lambda_0(\mathcal{M})$ we can define $W(g) = D(g)(p)$. It also suggests that a selection of point derivations, one for each $p \in \mathcal{M}$, might be used to create a derivation on $\Lambda_0(\mathcal{M})$.

There are some problems with this, however. Members of $\mathcal{F}_p^\infty(\mathcal{M})$ are not necessarily differentiable (or even defined) away from a vicinity of p, so every $\mathcal{F}_p^\infty(\mathcal{M})$ contains more functions than $\Lambda_0(\mathcal{M})$. There is a domain mismatch.

But we saw that if x is a coordinate map around p, a point derivation at p is defined by what it does to the coordinate functions x^i on an arbitrarily small neighborhood of p. Off this neighborhood x^i can be extended to a smooth function defined on all of $\mathcal{M}$, and D *is* defined on these extensions. So there is one and only one point derivation at p which agrees with W on all of $\Lambda_0(\mathcal{M})$.

Going the other way requires something extra. A selection of a point derivation W_p (that is, a tangent vector) for each $p \in \mathcal{M}$ must vary from place to place on $\mathcal{M}$ in a smooth way if the output values $W_p(g)(p)$ are to combine to produce a differentiable function (i.e. a function in $\Lambda_0(\mathcal{M})$) on $\mathcal{M}$ for all $g \in \Lambda_0(\mathcal{M})$.

We close this section with the following important point. Suppose $\lambda \in \mathcal{T}_1^0(\mathcal{M}) = \Lambda_1(\mathcal{M})$. At points in a coordinate neighborhood U_x it can be represented as $\lambda_j\, dx^j$ for certain smooth coefficient functions λ_j.

Any such λ can be used to create a map

$$Z\colon \mathcal{T}_0^1(\mathcal{M}) \to \mathcal{F}^\infty(\mathcal{M}) \quad \text{by} \quad Z(X)(p) = \lambda_p X_p.$$

Z is not only a real vector space homomorphism, but actually it is an **$\mathcal{F}^\infty(\mathcal{M})$-module homomorphism.**

$$(gZ)(X)(p) = g(p)\lambda_p\, X_p = \lambda_p\, g(p) X_p = Z(gX)(p).$$

In other words, $gZ(X) = Z(gX)$.

It is interesting to note that the converse also holds.

25.1. ***Theorem.*** *Suppose $Z\colon \mathcal{T}_0^1(\mathcal{M}) \to \mathcal{F}^\infty(\mathcal{M})$ is any $\mathcal{F}^\infty(\mathcal{M})$-module homomorphism. There is a unique covector field $\lambda \in \mathcal{T}_1^0(\mathcal{M})$ for which $Z(X) = \lambda\, X$ for every $X \in \mathcal{T}_0^1(\mathcal{M})$.*

Proof. We first note that if X and Y are any two members of $\mathcal{T}_0^1(\mathcal{M})$ and Z is an $\mathcal{F}^\infty(\mathcal{M})$-module homomorphism and if X and Y agree on *any* neighborhood of a point p then $Z(X)(p) = Z(Y)(p)$. To see this let V be an open neighborhood of agreement. Then there is a smooth function $f\colon \mathcal{M} \to [0,1]$ for which $supp(f) \subset V$ and $f(p) = 1$. But then

$$Z(X) = f(p)\, Z(X)(p) = Z(f\, X)(p) = Z(f\, Y)(p) = f(p)\, Z(Y)(p) = Z(Y)(p).$$

So Z only notices the behavior of a field X in arbitrarily small neighborhoods of a point p when it calculates $Z(X)(p)$.

We expand on this point below.

There is an atlas for $\mathcal{M}$ consisting of coordinate systems $x\colon U_x \to R_x$ with the following properties. (i) There are open sets $V_x \subset \overline{V}_x \subset W_x \subset \overline{W}_x \subset U_x$. (ii) The sets V_x also form an open cover of $\mathcal{M}$. (iii) There is a smooth function $f_x\colon \mathcal{M} \to [0,1]$ for which $V_x \subset f_x^{-1}(1)$ and $supp(f_x) \subset \overline{W}_x$.

Suppose X is a generic vector field from $\mathcal{T}_0^1(\mathcal{M})$.

So X can be represented everywhere within U_x as

$$X_q = X^i(x)_q \left.\frac{\partial}{\partial x^i}\right|_q \quad \text{where} \quad X^i(x)\frac{\partial}{\partial x^i} \in \mathcal{T}_0^1(U_x).$$

If we define vector field $\overline{X}$ to be $f_x^2 X$ on U_x and the zero tensor off U_x we have created a member of $\mathcal{T}_0^1(\mathcal{M})$ that agrees with X everywhere in V_x and is the zero field off $\overline{W}_x$. For notational convenience in the following calculation we declare $f_x(q)X^i(x)_q$ to be the number 0 and $f_x(q) \left.\frac{\partial}{\partial x^i}\right|_q$ to be the zero tensor whenever $f_x(q) = 0$, providing a meaning for the coefficients and the coordinate form of the tensor even when $q \notin U_x$. With that temporary convention we have

$$\overline{X} = f_x\, X^i(x)\, f_x\, \frac{\partial}{\partial x^i} \quad \text{everywhere in } \mathcal{M}.$$

Now suppose p is a point in V_x and $Z\colon \mathcal{T}_0^1(\mathcal{M}) \to \mathcal{F}^\infty(\mathcal{M})$ is an $\mathcal{F}^\infty(\mathcal{M})$-module homomorphism.

$$\begin{aligned} Z(X)(p) &= f_x^2 Z(X)(p) = Z(f_x^2\, X)(p) = Z\left(\overline{X}\right)(p) = Z\left(f_x\, X^i(x)\, f_x\, \frac{\partial}{\partial x^i}\right)(p) \\ &= f_x(p)\, X^i(x)(p)\, Z\left(f_x\, \frac{\partial}{\partial x^i}\right)(p) = X^i(x)(p)\, Z\left(f_x\, \frac{\partial}{\partial x^i}\right)(p). \end{aligned}$$

We now define covector λ_p to be $Z\left(f_x\, \frac{\partial}{\partial x^i}\right)(p)\, dx_p^i$.

Note that $\lambda_p\, X_p = Z(X)(p)$.

Since Z produces smooth functions λ defined in this way is a member of $\mathcal{T}_1^0(V_x)$. The only property used about f_x was its smoothness and the fact that it is 1 on a neighborhood of p and 0 outside an open set whose closure is contained in U_x. With the remark at the start of the proof, any f_x will produce the same λ_p.

The only question that remains is the extent to which all this depends on the coordinate system x.

Suppose $y\colon U_y \to R_y$ is another coordinate system around p. We may replace f_x and f_y with a similar smooth function g which is 1 on $f_x^{-1}(1) \cap f_y^{-1}(1)$ and zero off $\overline{W}_x \cap \overline{W}_y$.

Following the same procedure we create a covector

$$\begin{aligned} \mu_p &= Z\left(f_y\, \frac{\partial}{\partial y^i}\right)(p)\, dy_p^i = Z\left(g^2\, \frac{\partial}{\partial y^i}\right)(p)\, dy_p^i \\ &= Z\left(g\, \frac{\partial x^j}{\partial y^i}\, g\, \frac{\partial}{\partial x^j}\right)(p)\, \frac{\partial y^i}{\partial x^t}(p)\, dx_p^t = g(p)\, \frac{\partial x^j}{\partial y^i}(p) \frac{\partial y^i}{\partial x^t}(p)\, Z\left(g\, \frac{\partial}{\partial x^j}\right)(p)\, dx_p^t \\ &= \delta_t^j\, Z\left(g\, \frac{\partial}{\partial x^j}\right)(p)\, dx_p^t = Z\left(g\, \frac{\partial}{\partial x^j}\right)(p)\, dx_p^j = \lambda_p. \end{aligned}$$

Our conclusion is that the covector field λ can be defined at every point in the manifold by a process that is independent of coordinate system, and which represents the homomorphism Z. □

25.2. ***Corollary.*** *Suppose*

$$Z\colon \underbrace{\mathcal{T}^0_1(\mathcal{M})\times\cdots\times\mathcal{T}^0_1(\mathcal{M})}_{r\text{ covector field factors}}\times\underbrace{\mathcal{T}^1_0(\mathcal{M})\;\times\cdots\times\mathcal{T}^1_0(\mathcal{M})}_{s\text{ vector field factors}}\to\mathcal{F}^\infty(\mathcal{M})$$

is $\mathcal{F}^\infty(\mathcal{M})$-multilinear. By this we mean it is an $\mathcal{F}^\infty(\mathcal{M})$-module homomorphism in each "slot" separately. There is a unique tensor field $T\in\mathcal{T}^r_s(\mathcal{M})$ for which

$$Z(\theta_1,\dots,\theta_r,X_1,\dots,X_s)=T(\theta_1,\dots,\theta_r,X_1,\dots,X_s)$$

for all choices of covector fields $\theta_i\in\mathcal{T}^0_1(\mathcal{M})$ and vector fields $X_i\in\mathcal{T}^1_0(\mathcal{M})$.

Proof. The proof is directly analogous to the last. This time examine $f_x^{2(r+s)}Z$ and distribute copies of f_x to all factors to establish the local nature of Z. □

26. Tensor Operations in a Coordinate Patch

We suppose $x\colon U_x\to R_x$ is a coordinate map for manifold $\mathcal{M}$. All functions and tensors on submanifold U_x are assumed to be restrictions to U_x of similar smooth objects defined on $\mathcal{M}$.

A member $T\in\mathcal{T}^r_s(U_x)$ is of the form

$$T^{i_1,\dots,i_r}_{j_1,\dots,j_s}(x)\,\frac{\partial}{\partial x^{i_1}}\otimes\cdots\otimes\frac{\partial}{\partial x^{i_r}}\otimes dx^{j_1}\otimes\cdots\otimes dx^{j_s}$$

for smooth functions $T^{i_1,\dots,i_r}_{j_1,\dots,j_s}(x)$ defined on U_x.

We write $T^{i_1,\dots,i_r}_{j_1,\dots,j_s}$ for $T^{i_1,\dots,i_r}_{j_1,\dots,j_s}(x)$ when coordinate map x is fixed.

If t is the identity map on $R_x\subset\mathbb{R}^n$ this can be pushed via x_* and $x^{-1\,*}$ to

$$T^{i_1,\dots,i_r}_{j_1,\dots,j_s}\circ x\;\frac{\partial}{\partial t^{i_1}}\otimes\cdots\otimes\frac{\partial}{\partial t^{i_r}}\otimes dt^{j_1}\otimes\cdots\otimes dt^{j_s}\in\mathcal{T}^r_s(R_x)$$

and calculations carried out in $\mathbb{R}^n$ if you prefer.

For $S,T\in\mathcal{T}^1_0(U_x)$ the tensor product $P=S\otimes T$ of S and T is given by

$$P=P^{i_1,i_2}\,\frac{\partial}{\partial x^{i_1}}\otimes\frac{\partial}{\partial x^{i_2}}=S^{i_1}T^{i_2}\,\frac{\partial}{\partial x^{i_1}}\otimes\frac{\partial}{\partial x^{i_2}}\in\mathcal{T}^2_0(U_x).$$

Similarly, for $\sigma,\tau\in\mathcal{T}^0_1(U_x)$ the tensor product $\theta=\sigma\otimes\tau$ of σ and τ is given by

$$\theta=\theta_{j_1,j_2}\,dx^{j_1}\otimes dx^{j_2}=\sigma_{j_1}\tau_{j_2}\,dx^{j_1}\otimes dx^{j_2}\in\mathcal{T}^0_2(U_x).$$

The mixed tensor product $M=T\otimes\tau\in\mathcal{T}^1_1(U_x)$ is

$$M=M^i_j\,\frac{\partial}{\partial x^i}\otimes dx^j=T^i\tau_j\,\frac{\partial}{\partial x^i}\otimes dx^j.$$

One contravariant index can be contracted against a covariant index in a "trace" operation given by number field

$$C^i_j(M) = M^i_i = T^i\tau_i.$$

Applied to a more general mixed tensor, the C operator specifies the index symbols that are to be replaced by an otherwise unused generic index and summed over. This reduces the degree of the original tensor by two: one covariant and one contravariant index disappear.

Now we come to wedge product.

Let $\mathcal{P}_L$ denote the set of permutations on the first L positive integers.

Suppose $\theta \in \mathcal{T}^0_L(U_x),\ \sigma \in \mathcal{T}^0_s(U_x)$ and $\tau \in \mathcal{T}^0_t(U_x)$.

$$\begin{aligned} Alt(\theta) &= \frac{1}{L!}\sum_{Q\in\mathcal{P}_L} sgn(Q)\,\theta_{i_1,\dots,i_L}\,dx^{i_{Q(1)}}\otimes\cdots\otimes dx^{i_{Q(L)}} \\ &= \frac{1}{L!}\sum_{Q\in\mathcal{P}_L} sgn(Q)\,\theta_{i_{Q(1)},\dots,i_{Q(L)}}\,dx^{i_1}\otimes\cdots\otimes dx^{i_L}. \end{aligned}$$

(Sum on an index pair i_m and $i_{Q(q)}$ whenever $m = Q(q)$.)

and

$$\begin{aligned} Sym(\theta) &= \frac{1}{L!}\sum_{Q\in\mathcal{P}_L} \theta_{i_1,\dots,i_L}\,dx^{i_{Q(1)}}\otimes\cdots\otimes dx^{i_{Q(L)}} \\ &= \frac{1}{L!}\sum_{Q\in\mathcal{P}_L} \theta_{i_{Q(1)},\dots,i_{Q(L)}}\,dx^{i_1}\otimes\cdots\otimes dx^{i_L}. \end{aligned}$$

(Sum on an index pair i_m and $i_{Q(q)}$ whenever $m = Q(q)$.)

$Alt(\theta) = \theta$ if and only if θ is alternating, while $Sym(\theta) = \theta$ if and only if θ is symmetric.

In some applications the coefficients of $Alt(\theta)$ and $Sym(\theta)$ are indicated in coordinates by

$$Alt(\theta) = \theta_{[i_1,\dots,i_L]}\,dx^{i_1}\otimes\cdots\otimes dx^{i_L} \quad\text{and}\quad Sym(\theta) = \theta_{(i_1,\dots,i_L)}\,dx^{i_1}\otimes\cdots\otimes dx^{i_L}.$$

For $f \in \mathcal{F}^\infty(U_x)$ and $\theta \in \mathcal{T}^0_L(U_x)$ we define $f\wedge\theta = \theta\wedge f = f\,\theta$.

Define **wedge product** between s-forms and t-forms by

$$\sigma\wedge\tau = \frac{(s+t)!}{s!\ t!}\,Alt(\sigma\otimes\tau).$$

From the standpoint of calculation this can be impossibly arduous in view of the fact that $\mathcal{P}_L$ has $L!$ members. However, it is not completely unmanageable when L is 4 or less, as it usually is.

Suppose now that $\mathcal{M}$ is a 3-manifold and $\sigma = 4\,dx^1 + 8\,dx^2$ and $\tau = 6\,dx^1 + 5\,dx^3$.

$\mathcal{P}_2$ contains two permutations only: the identity and the permutation that switches 1 and 2.

$$\begin{aligned}
\gamma = \sigma \wedge \tau &= \frac{(2)!}{1!\,1!}\, Alt(\sigma \otimes \tau) = 2\, Alt(\sigma \otimes \tau) \\
&= 2\, Alt(\, 24\, dx^1 \otimes dx^1 + 20\, dx^1 \otimes dx^3 + 48\, dx^2 \otimes dx^1 + 40\, dx^2 \otimes dx^3\,) \\
&= 24\, dx^1 \otimes dx^1 + 20\, dx^1 \otimes dx^3 + 48\, dx^2 \otimes dx^1 + 40\, dx^2 \otimes dx^3 \\
&\qquad - (24\, dx^1 \otimes dx^1 + 20\, dx^3 \otimes dx^1 + 48\, dx^1 \otimes dx^2 + 40\, dx^3 \otimes dx^2) \\
&= 20\, (dx^1 \otimes dx^3 - dx^3 \otimes dx^1) - 48\, (dx^1 \otimes dx^2 - dx^2 \otimes dx^1) \\
&\qquad\qquad + 40\, (dx^2 \otimes dx^3 - dx^3 \otimes dx^2) \\
&= 20\, dx^1 \wedge dx^3 - 48\, dx^1 \wedge dx^2 + 40\, dx^2 \wedge dx^3.
\end{aligned}$$

Finally, we define the interior product.

If S is a tangent vector field and f is a smooth function, a member of $\Lambda_0(U_x) = \mathcal{F}^\infty(U_x)$ we define the **"angle" operation**, also called the **interior product of S on f**, by $S \lrcorner f = 0$.

More generally, for $\gamma \in \Lambda_r(U_x)$ we define $S \lrcorner \gamma$, the **interior product of S with $\boldsymbol{\gamma}$**, to be the contraction of the tensor product of S against γ in the first index of γ.

So $S \lrcorner \gamma \in \Lambda_{r-1}(U_x)$ for each $r > 0$.

If $r = 1$ this is nothing more than the trace, the evaluation of S at γ.

Suppose $r = 2$ and

$$\gamma = 20\, dx^1 \wedge dx^3 - 48\, dx^1 \wedge dx^2 + 40\, dx^2 \wedge dx^3$$

$$\text{and} \quad S = -9\, \frac{\partial}{\partial x^1} + \frac{\partial}{\partial x^2} - 20\, \frac{\partial}{\partial x^3}.$$

Distributing, using the expanded representation of γ, gives

$$S \lrcorner \gamma = 448\, dx^1 + 1232\, dx^2 - 140\, dx^3.$$

A faster route to the same result is to use the identity

$$S \lrcorner\, (\sigma \wedge \tau) = (\, S \lrcorner \sigma\,) \wedge \tau + (-1)^r\, \sigma \wedge (\, S \lrcorner \tau\,).$$

27. Pullback on Covariant Tensors

We have seen before that if $H \colon \mathcal{M} \to \mathcal{N}$ is smooth we can pullback a covector field $\omega \in \mathcal{T}_1\,(\mathcal{N})$ to a covector field $H^*(\omega) \in \mathcal{T}_1\,(\mathcal{M})$.

For each $H(p) \in \mathcal{N}$ the tensor $\omega_{H(p)}$ is $df_{H(p)}$ for real valued smooth function f defined in a neighborhood of $H(p)$.

The tensor $H^*(\omega)_p \in \mathcal{M}_p^*$ is just $d(f \circ H)_p$.

For each $p \in \mathcal{M}$ there are coordinate neighborhoods U_y around $H(p)$ and U_x around p for which $H^{-1}(U_y) \subset U_x$.

For points q around $H(p)$ the tensor $\omega(q) = \omega_i(q)\, dy_q^i$ for certain smooth functions $\omega_i(q)$.

Then for points $p \in U_x$ we found that

$$H^*(\omega)_p = \omega_i(H(p))\, d(y^i \circ H)_p = \omega_i(H(p))\, \frac{\partial(y^i \circ H)}{\partial x^j}\, dx^j_p.$$

The procedure for more general covariant tensors is essentially identical. For each $p \in \mathcal{M}$ and $\omega \in \mathcal{T}_L(\mathcal{N})$ there is a representation, valid for q near $H(p)$,

$$\omega(q) = \omega_{i_1,\dots,i_L}(q)\, dy^{i_1}_q \otimes \cdots \otimes dy^{i_L}_q.$$

We then have

$$\begin{aligned}
H^*(\omega)_p &= \omega_{i_1,\dots,i_L}(H(p))\, H^*(dy^{i_1})_p \otimes \cdots \otimes H^*(dy^{i_L})_p \\
&= \omega_{i_1,\dots,i_L}(H(p))\, d(y^{i_1} \circ H)_p \otimes \cdots \otimes d(y^{i_L} \circ H)_p \\
&= \omega_{i_1,\dots,i_L}(H(p))\, \frac{\partial(y^{i_1} \circ H)}{\partial x^{j_1}}\, dx^{j_1}_p \otimes \cdots \otimes \frac{\partial(y^{i_L} \circ H)}{\partial x^{j_L}}\, dx^{j_L}_p \\
&= \left(\omega_{i_1,\dots,i_L}(H(p))\, \frac{\partial(y^{i_1} \circ H)}{\partial x^{j_1}} \cdots \frac{\partial(y^{i_L} \circ H)}{\partial x^{j_L}}\right) dx^{j_1}_p \otimes \cdots \otimes dx^{j_L}_p
\end{aligned}$$

There is a similar formula for the alternating tensors $\omega \in \Lambda_L(\mathcal{N})$. There, in a neighborhood of $H(p)$ we have

$$\omega(q) = \omega_{i_1,\dots,i_L}(q)\, dy^{i_1}_q \wedge \cdots \wedge dy^{i_L}_q. \quad \text{(Sum on increasing indices only.)}$$

Below, $H^*(\omega)_p$ is given as a sum where $i_1, \dots, i_L$ are increasing indices but $j_1, \dots, j_L$ can take any values:

$$H^*(\omega)_p = \left(\omega_{i_1,\dots,i_L}(H(p))\, \frac{\partial(y^{i_1} \circ H)}{\partial x^{j_1}} \cdots \frac{\partial(y^{i_L} \circ H)}{\partial x^{j_L}}\right) dx^{j_1}_p \wedge \cdots \wedge dx^{j_L}_p.$$

If any of the values on the list $j_1, \dots, j_L$ are duplicated in a term, the wedge product is 0, and any two lists that permute the same $j_1, \dots, j_L$ values can be combined to a single term with increasing $j_1, \dots, j_L$ by introducing a minus sign where appropriate.

The case of n-forms when the dimension of $\mathcal{N}$ and $\mathcal{M}$ are also n will be of special interest. In that situation the tensor spaces are 1 dimensional at each point, so the formula becomes

$$H^*(\omega)_p = \omega(H(p)) \left(\frac{\partial(y^1 \circ H)}{\partial x^{j_1}} \cdots \frac{\partial(y^n \circ H)}{\partial x^{j_n}}\right) dx^{j_1}_p \wedge \cdots \wedge dx^{j_n}_p.$$

Upon introducing the factor $sgn(P)$ when $j_1, \dots, j_n$ is an odd permutation of $1, \dots, n$ we see that

$$H^*(\omega)_p = \omega(H(p))\, \det\left(\frac{d(y \circ H)}{dx}\right) dx^1_p \wedge \cdots \wedge dx^n_p.$$

28. The Metric

A **metric tensor on $\mathcal{M}$** is a symmetric nondegenerate member of $\mathcal{T}_2(\mathcal{M})$. In this generality they are often called **semi-Riemannian metrics**.

Semi-Riemannian metrics are (not necessarily positive definite) inner products on the tangent space on $\mathcal{M}_q$ at each point q on the manifold.

A positive definite metric tensor is called a **Riemannian metric**.

In a coordinate patch U_x a metric tensor g has the form

$$g = g_{i,j}\, dx^i \otimes dx^j.$$

The coordinates $g_{i,j}$ are functions depending on the point p of the manifold and on the coordinate map x, but the notation $(g_p)_{i,j}(x)$ is too cumbersome for daily use, so it is expected that the user understands that these functions are to be evaluated at p for each x and in these coordinates g is to be applied to pairs of tensor fields X and Y also evaluated at p and represented in coordinates using x.

In this patch, symmetry amounts to the condition $g_{i,j} = g_{j,i}$ for all i, j and this means $g_p(X_p, Y_p) = g_p(Y_p, X_p)$ for all pairs of vector fields X and Y and all $p \in \mathcal{M}$.

Nondegeneracy is defined "pointwise" using vector fields. g is nondegenerate if for each point p in each coordinate patch the only way that $g_p(X_p, Y_p)$ can equal 0 for all possible Y_p is if X_p itself is 0.

Positive definiteness is the condition that $g_p(X_p, X_p) \geq 0$ for all vector fields X and all $p \in \mathcal{M}$.

At a point p (*not* on a neighborhood, just at p) every semi-Riemannian metric can be diagonalized, to yield a representation of the form

$$g = \sum_{i=1}^{n} g_{i,i}\, dx^i \otimes dx^i \quad \text{where each} \quad g_{i,i} = \pm 1 \quad \text{at } p.$$

Continuity and non-degeneracy of g on $\mathcal{M}$ imply that ***if $\mathcal{M}$ is connected*** the number of ones and minus ones in any such representation cannot change on $\mathcal{M}$.

The metric is said to be of type (k, l) if there are k ones and l minus ones in a representation of this kind. Note $k + l = n$ where n is the dimension of $\mathcal{M}$.

We will be interested in two types of metrics: the positive definite ones which have type $(n, 0)$ and which generalize the ordinary Euclidean metric, and metrics of type $(3, 1)$ which are semi-Riemannian and generalize the Minkowski metric of space-time.

There is a commonly used "raising or lowering index" operation obtained through the services of an inner product at each point in U_x.

We specify metric tensor $g = g_{i,j}\, dx^i \otimes dx^j \in \mathcal{T}_2(U_x)$.

The matrix obtained using the coefficients of g in this basis is symmetric and invertible. Let $g^{i,j}$ denote the entries of the inverse to this matrix.

The conjugate metric tensor, which we will denote $g^\star$, is defined by $g^\star = g^{i,j}\, \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}$ and is also symmetric and nondegenerate and defines an inner product on $\mathcal{M}_q^*$ for each $q \in U_x$.

To "lower" an index on vector field $T = T^i \frac{\partial}{\partial x^i}$ take the tensor product $g \otimes T$ and contract, creating

$$\flat(T) = \flat(T)_j \, dx^j = g_{i,j} T^i \, dx^j.$$

To "raise" the index on $\theta = \theta_j \, dx^j$ take the tensor product $g^\star \otimes \theta$ and contract, turning θ into

$$\sharp(\theta) = \sharp(\theta)^i \, \frac{\partial}{\partial x^i} = g^{i,j} \theta_j \, \frac{\partial}{\partial x^i}.$$

These processes are inverse to each other and the sharp and flat symbols are intended to suggest turning a flat object, a covector, into a sharp object, a vector. One "sharpens" a covector and "flattens" a tangent vector.

It is unfortunate that standard practice is to think of these vectors and covectors as "the same" and leave off the notational distinction. You deduce what has been done to T or θ by inspecting the location of the index, high or low, in a representation in a basis.

Here is an example. Let $\mathcal{M}$ be a 4-manifold and suppose on the patch U_x

$$g = g_{i,j} \, dx^i \otimes dx^j = -\, dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3 + dx^4 \otimes dx^4.$$

Here we are *supposing* a situation that is not at all generic: we suppose it is possible to diagonalize g on a whole neighborhood, not just at one point.

In this case the matrix of coefficients is its own inverse and

$$\begin{aligned} g^\star &= g^{i,j} \, \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \\ &= -\, \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^1} + \frac{\partial}{\partial x^2} \otimes \frac{\partial}{\partial x^2} + \frac{\partial}{\partial x^3} \otimes \frac{\partial}{\partial x^3} + \frac{\partial}{\partial x^4} \otimes \frac{\partial}{\partial x^4}. \end{aligned}$$

Given vector $T = 8 \frac{\partial}{\partial x^1} + 4 \frac{\partial}{\partial x^2} - 2 \frac{\partial}{\partial x^3} + 2 \frac{\partial}{\partial x^4}$ we can flatten T to $-8 \, dx^1 + 4 \, dx^2 - 2 \, dx^3 + 2 \, dx^4$ and then sharpen it again with the reverse operation.

T has constant inner product $g(T,T) = -64 + 16 + 4 + 4 = -40$.

29. Integration on Manifolds

There is no natural way to define the integral of a real-valued function defined on a smooth manifold $\mathcal{M}$ that matches the properties of ordinary Riemann integration: additional structure is required.

We will use concepts from $\mathbb{R}^n$ and coordinates to provide this structure, carving out pieces of our manifold and integrating forms on these pieces, in much the same way as the Riemann integral is initially defined on closed intervals or n-dimensional rectangles.

We deal with a trivial special case first. We define $\mathbf{[0,1]^0}$ to be the real number 0 and a **singular 0-cube in manifold $\mathcal{M}$** is the selection $h(0)$ of a point in $\mathcal{M}$.

Now if f is any 0-form on $\mathcal{M}$ (that is, a smooth function) we define

$$\int_h \lambda = f(h(0)).$$

For $k > 0$ a smooth function $h\colon [0,1]^k \to \mathcal{M}$ will be called a **singular k-cube in $\mathcal{M}$** and the inclusion function $I^k\colon [0,1]^k \to \mathbb{R}^k$ is the simplest case, called the **standard k-cube**.

When dealing with issues of integration in the manifold $\mathbb{R}^k$, we will let x stand for the identity coordinate system there. So x agrees with (i.e. *is*) I^k on $[0,1]^k$.

Since $\Lambda_k([0,1]^k)$ has dimension 1 at each point any k-form $\lambda \in \Lambda_k([0,1]^k)$ can be written as

$$f\, dx^1 \wedge \cdots \wedge dx^k$$

where f is smooth on $[0,1]^k$.

We define the integral $\int_{I^k} \boldsymbol{\lambda}$ of the k-form λ on the standard k-cube by the usual iterated Riemann integral

$$\int_{I^k} \lambda = \int_0^1 \cdots \int_0^1 f(x)\, dx^1 \cdots dx^k.$$

If $h\colon [0,1]^k \to \mathcal{M}$ is a singular k-cube in n-manifold $\mathcal{M}$ then we can define the pullback

$$h^*\colon \Lambda_k(\mathcal{M}) \to \Lambda_k([0,1]^k).$$

If $p \in [0,1]^k$ then $h(p)$ is in some coordinate neighborhood U_y in $\mathcal{M}$. So if $\lambda \in \Lambda_k(\mathcal{M})$ it can be represented near $h(p)$ as

$$\lambda_{i_1,\ldots,i_k}\ dy^{i_1} \wedge \cdots \wedge dy^{i_k}$$

where we sum here on increasing indices only.

Then $h^*(\lambda)$ is given near p by

$$(\lambda_{i_1,\ldots,i_k} \circ h)\ d(y^{i_1} \circ h) \wedge \cdots \wedge d(y^{i_k} \circ h).$$

The terms in this sum are all multiples of $dx^1 \wedge \cdots \wedge dx^k$ and so, as before, $h^*(\lambda)$ is $f\, dx^1 \wedge \cdots \wedge dx^k$ for some smooth function f defined on all of $[0,1]^k$.

The function f has a particular form near p. Defining the function

$$y^{i_1,\ldots,i_k} = (y^{i_1}, \ldots, y^{i_k})$$

from an open neighborhood of $h(p) \in \mathcal{M}$ to $\mathbb{R}^k$ and using the properties of wedge product and $d(y^{i_t} \circ h) = \frac{\partial (y^{i_t} \circ h)}{dx^j} dx^j$ we see that a factor of the Jacobian determinant appears:

$$d(y^{i_1} \circ h) \wedge \cdots \wedge d(y^{i_k} \circ h) = \det\left(\frac{d(y^{i_1,\ldots,i_k} \circ h)}{dx}\right) dx^1 \wedge \cdots \wedge dx^k.$$

If $\mathcal{M}$ is a k-manifold there is just one increasing index sequence and that factor is simply $\det\left(\frac{d(y \circ h)}{dx}\right)$. In any case, near p we have the function f as

$$\begin{aligned} h^*(\lambda) =& (\lambda_{i_1,\ldots,i_k} \circ h)\ d(y^{i_1} \circ h) \wedge \cdots \wedge d(y^{i_k} \circ h) \\ =& (\lambda_{i_1,\ldots,i_k} \circ h)\ \det\left(\frac{d(y^{i_1,\ldots,i_k} \circ h)}{dx}\right) dx^1 \wedge \cdots \wedge dx^k \\ =& f(x)\, dx^1 \wedge \cdots \wedge dx^k. \end{aligned}$$

And if $\mathcal{M}$ is a k-manifold then

$$f(x) = (\lambda \circ h)\ \det\left(\frac{d(y \circ h)}{dx}\right).$$

We now define the **integral of the k-form $\boldsymbol{\lambda}$ over the singular $\boldsymbol{k}$-cube $\boldsymbol{h}$** by

$$\int_h \lambda = \int_{I^k} h^*(\lambda) = \int_0^1 \cdots \int_0^1 f(x)\, dx^1 \cdots dx^k.$$

At this point we have identified a way of "carving out" a piece of a manifold (the image of a singular k-cube) and integrating k-forms on that piece. It is obvious by our definition that

$$\int_h (\lambda + c\mu) = \int_h \lambda + c \int_h \mu$$

for singular k-cube h and k-forms λ and μ and real c so our definition of integral on manifolds also has the vital linearity property possessed by the Riemann Integral.

Now suppose $s\colon [0,1]^k \to [0,1]^k$ is a diffeomorphism. This implies that $\frac{ds}{dx}$ is invertible on $[0,1]^k$ and so (since $[0,1]^k$ is connected) $\det\left(\frac{ds}{dx}\right)$ is nonzero with constant sign. The diffeomorphism is said to be **orientation preserving** if this determinant is positive everywhere, and **orientation reversing** if the determinant is negative everywhere.

The function $g = h \circ s$ is called a **reparametrization** of the singular k-cube h and g is also a singular k-cube.

We want to know the relationship between $\int_h \lambda$ and $\int_g \lambda = \int_{h\circ s} \lambda$ for k-form λ.

If $h^*(\lambda) = f(x)\, dx^1 \wedge \cdots \wedge dx^k$ then

$$g^*(\lambda) = (h \circ s)^*(\lambda) = s^*(h^*(\lambda)) = s^*(f(x)\, dx^1 \wedge \cdots \wedge dx^k).$$

But that means

$$\begin{aligned}
\int_g \lambda = \int_{h\circ s} \lambda = \int_s h^*(\lambda) &= \int_s f(x)\, dx^1 \wedge \cdots \wedge dx^k \\
&= \int_0^1 \cdots \int_0^1 (f \circ s)(x)\ \det\left(\frac{ds}{dx}\right)\, dx^1 \cdots dx^k \\
&= \pm \int_0^1 \cdots \int_0^1 f(x)\, dx^1 \cdots dx^k = \pm \int_h \lambda
\end{aligned}$$

by the ordinary change-of-variable formula for Riemann integrals, where the "plus or minus" choice is determined by the sign of the determinant.

In other words, any singular* k*-cube that is a reparameterization of a given singular* k*-cube will yield the same integral when applied to any* k*-form, up to a sign determined by the "orientation" of the reparameterizing diffeomorphism.

For Riemann integrals the Fundamental Theorem of Calculus provides a tool to actually calculate integrals, relating the integral of the derivative of a continuously differentiable function on a closed interval with values of the function itself on the *boundary* of that interval, with a minus sign associated with one of the two boundary points.

There is an analogue of this result for our integral, contained in **Stokes' Theorem**, which we will prove after we have created a way of differentiating forms on a manifold.

30. Singular k-Chains and Their Integrals

In this section we set the stage for Stokes' Theorem by forming objects called **singular k-chains** and the **boundaries** of these.

First, is the easy part.

A **singular k-chain** is any finite formal sum of the form

$$\mathcal{C} = m_1\, h_1 + \cdots + m_t\, h_t$$

where the m_i are all integers and the h_i are all singular k-cubes into the same manifold $\mathcal{M}$.

We declare that the order of these terms doesn't matter[19] and $m_1\, h + m_2\, h = (m_1 + m_2)\, h$ for integers m_1, m_2 and singular k-cube h.

Singular k-chains are added by combining the coefficients on identical k-cube summands. The formal additive identity will be denoted 0, which may be realized as $0\, h$ for any singular k-cube h.

We make these formal sums into a module over the integers in the obvious way.

We define[20] the **integral of k-form $\boldsymbol{\lambda}$ on singular k-chain $\boldsymbol{\mathcal{C}}$** by

$$\int_{\mathcal{C}} \lambda = m_1 \int_{h_1} \lambda + \cdots + m_t \int_{h_t} \lambda.$$

We now define the boundary of singular k-chains.

Recall that a singular 0-cube is defined to be a selection of a point in a manifold. We define the **boundary $\boldsymbol{\partial h}$** of a singular 0-cube h to be the number 1.

A singular 1-cube is a smooth map $h\colon [0,1] \to \mathcal{M}$. We define the **boundary $\boldsymbol{\partial h}$** of a singular 1-cube h to be the singular 0-chain $h(1) - h(0)$ where we interpret $h(1)$ to denote the 0-cube which "selects" $h(1) \in \mathcal{M}$ and interpret $h(0)$ similarly.

The **boundary** of the standard k-cube $I^k\colon [0,1]^k \to \mathbb{R}^k$ is a $(k-1)$-chain composed of $2k$ different singular $(k-1)$-cubes, each of which corresponds to one of the $2k$ "faces" of the cube.

Following Spivak's notation, we define for $k > 1$ and $i = 1, \ldots, k$ and $j = 0, 1$

$$\boldsymbol{I^k_{i,j}\colon [0,1]^{k-1} \to [0,1]^k} \quad \text{via} \quad (x_1, \ldots, x_{k-1}) \to (x_1, \ldots, x_{i-1}, j, x_i, \ldots, x_{k-1}).$$

Then the **boundary of the standard k-cube** is given by

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

[19] The formal definition of the equality of two formal sums of singular k-cubes in terms of equivalence classes of all the formal sums we intend to be equal is left for the energetic reader.

[20] It may be obvious that this definition does not depend on the representative of $\mathcal{C}$ used to define it, but that fact does require a tedious little proof.

and more generally for any singular k-cube $h\colon [0,1]^k \to \mathcal{M}$

$$\boldsymbol{\partial h} = \sum_{i=1}^{k} (-1)^i h \circ I^k_{i,0} + (-1)^{i+1} h \circ I^k_{i,1}.$$

And finally, if $\mathcal{C} = m_1 h_1 + \cdots + m_t h_t$ is any singular k-chain we define the **boundary of $\mathcal{C}$** to be the $(k-1)$-chain

$$\boldsymbol{\partial \mathcal{C} = m_1 \partial h_1 + \cdots + m_t \partial h_t.}$$

We are going to use these boundaries when we prove Stokes' Theorem, but for now we just want to prove a fact that will be important in applications of that theorem: that $\partial\partial\mathcal{C} = \partial^2\mathcal{C} = 0$. A boundary has no boundary.

If we can prove this fact when applied to the standard singular k-cube the result will follow for more general singular k-chains with minor adaptation. And we restrict attention to k at least 1.

Considering the case $k = 1$ first we have $\partial I^1 = 1 - 0$ where in this case $1 - 0$ represents the difference of two singular 0-cubes. So $\partial^2 I^1 = \partial 1 - \partial 0 = 1 - 1 = 0$.

So what happens when $k = 2$? Hoping to be forgiven for rather loose function notation, we calculate

$$\begin{aligned}\partial^2 I^2 =& \partial\,(\ (-1)(0,x) + (1,x) + (x,0) + (-1)(x,1)\ )\\ =& (-1)\,[\ (-1)(0,0) + (1,0) + (0,0) + (-1)(0,1)\ ]\\ &\qquad + [\ (-1)(0,1) + (1,1) + (1,0) + (-1)(1,1)\ ] = 0.\end{aligned}$$

For larger k we have

$$\begin{aligned}\partial^2 I^k =& \sum_{i=1}^{k} (-1)^i \partial I^k_{i,0} + (-1)^{i+1} \partial I^k_{i,1}\\ =& \sum_{i=1}^{k} \sum_{L=1}^{k-1} (-1)^{i+L} I^k_{i,0} \circ I^{k-1}_{L,0} + (-1)^{i+L+1} I^k_{i,0} \circ I^{k-1}_{L,1}\\ &\qquad + (-1)^{i+1+L} I^k_{i,1} \circ I^{k-1}_{L,0} + (-1)^{i+1+L+1} I^k_{i,1} \circ I^{k-1}_{L,1}\end{aligned}$$

We have four double sums each consisting of $k(k-1)$ terms, a huge sum of singular $(k-2)$-cubes. Each term in the sum is a function whose values are of the form

$$(x_1, \ldots, x_{s-1}, A_{\text{first}}, x_s, \ldots, x_{t-1}, B_{\text{second}}, x_t, \ldots, x_{k-2})$$

or

$$(x_1, \ldots, x_{s-1}, A_{\text{second}}, x_s, \ldots, x_{t-1}, B_{\text{first}}, x_t, \ldots, x_{k-2})$$

where the A and B values are either 0 or 1 and the subscript indicates which of the two summation indices (i first, L second) put that particular value there.

As indicated, each one of these terms for specific A and B will occur twice, **first** when the summation on i puts the appropriate value at A, and then the sum on L must put the appropriate value at B. And **then a second time** when the summation on i puts the appropriate value at B, and then the sum on L must put the appropriate value at A.

In the first case $i = s$ and $L = t$ but in the second case $i = t + 1$ and $L = s$.

Each of these two terms has a sign attached, and the signs are opposite, so the entire sum is 0 and the result is proved.

31. Example Integrals

References

[1] Michael Spivak, *A Comprehensive Introduction to Differential Geometry Volume 1, Third Edition* 1999: Publish or Perish Press.

[2] John Lee, *Introduction to Smooth Manifolds 2nd Ed.* 2012: Springer.

[3] William Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry Revised 2nd Ed.* 2003: Academic Press.

[4] Serge Lang, *Introduction to Differentiable Manifolds 2nd Ed.* 2002: Springer.

[5] Michael Spivak, *Calculus on Manifolds* 1965: Benjamin.

[6] Garrett Birkhoff and Gian-Carlo Rota, *Ordinary Differential Equations 4th Ed.* 1989: John Wiley & Sons.

[7] Loomis and Sternberg, *Advanced Calculus Revised Ed.* 2014: World Scientific.

Index