Supplement for Mathematics Texts

Stefan Lukits

November 7, 2017

Contents

1 Bernhard Schutz: Geometrical Methods of Mathematical Physics

1.1 Some Basic Mathematics

On page 15, show that $d''(x, y)$ is not a norm.

1.2 Differentiable Manifolds and Tensors

On page 28, cover the interior of the annulus by a single coordinate patch.

1.2.1 Vectors and Vector Fields

Explain (2.2) on page 32. Schutz' λ is confusing to me. It appears to be a function from a closed interval $[a, b] \subset \mathbb{R}$ into the manifold M (a curve). f is $C^\infty(M)$, so $f : U \to \mathbb{R}$. g is just a coordinate map, what Jeff Lee calls x, and goes from U to $\mathbb{R}^n$. $g : U \to \mathbb{R}$. This makes no sense to me. What the hell is

$$\frac{dg}{d\lambda} \tag{1}$$

if g is defined on U and goes into $\mathbb{R}^n$?

Answer: you are mistaking figure 2.9 to apply to section 2.7 instead of section 2.6. g is not a coordinate map in section 2.7. It looks like the x^i determine the curve. In Jeff Lee's notation, let $\gamma : (a, b) \to V$ ((V, y) is the usual open neighbourhood in M plus coordinate map pair). Then Schutz'

$$(x^1(\lambda), \ldots, x^n(\lambda)) \tag{2}$$

corresponds to Lee's $(y \circ \gamma)(\lambda)$ in $\mathbb{R}^n$. Schutz' $f : \mathbb{R}^n \to \mathbb{R}$ corresponds to

$$\hat{f} : \xi \in \mathbb{R}^n \to f(y^{-1}(\xi)) \tag{3}$$

in Lee. I derived (2.2) in Schutz with a complicated translation table between Lee and Schutz on page 2256 in Schmierbuch. Write g in Schutz instead as $\hat{f} \circ x$, where $\hat{f}$ is as defined in (3) and x is as defined in (2), unfortunately mixing Schutz and Lee, but (2.2) pretty handily follows.

Here is an example for the chain rule and multivariable differentiation. Let

$$\begin{aligned} f(x) &= \left(x^2, \tfrac{x}{2}\right) \\ g(w, z) &= 3w - z^2 \end{aligned} \tag{4}$$

Then $(g \circ f)(x) = \frac{11}{4}x^2$ and therefore $(g \circ f)'(x) = \frac{11}{2}x$. According to the chain rule

$$\frac{d(g \circ f)}{dx} = \sum \frac{df^i}{dx} \frac{\partial(g \circ f)}{\partial f^i} = 2x \cdot 3 + \frac{1}{2} \cdot \left(-2 \cdot \frac{x}{2}\right) = \frac{11}{2}x \tag{5}$$

This is what is behind Schutz' equation (2.2) on page 32.

1.2.2 Exercise 2.1

On page 44.

2 Michael K. Murray and John W. Rice: Differential Geometry and Statistics

Note that there is an informative article on exponential families and categorical distributions in Wikipedia. It suggests as parameters for the categorical distribution

$$\left(\ln \frac{p_1}{p_k}, \ldots, \ln \frac{p_{k-1}}{p_k}, 0\right)^{\intercal}. \tag{6}$$

I wonder if the last parameter is necessary, since it's always zero.

Let P be the parametric family of normal distributions. Let $q = \mathcal{N}(0,1)$ (the origin) and $p = \mathcal{N}(7,9)$. The density for the normal distribution is

$$p(\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{7}$$

Expressed as an exponential family, this is

$$p(\theta^1,\theta^2) = \exp\left(\theta^2 x^2 + \theta^1 x^1 - K(\theta)\right) \tag{8}$$

x^2 is a second random variable (using differential geometry's habit of indexing vector components on the top right), but in the case of the normal distribution it is the first random variable squared, which makes for a hell of notational confusion.

For q and p as above, the parameters are $(-0.5, 0)$ and $(-1/18, 7/9)$, respectively. The origin q is not at $(0,0)$, which also creates confusion. The vector $\vec{w}$ associated with p, for example, so that

$$q \oplus \vec{w} = p \tag{9}$$

where $q \oplus \vec{w} = exp(\vec{w}x)q(x)$ (x is (x^1, x^2)), is

$$\vec{w}^\intercal = \left(\frac{4}{9}, \frac{7}{9}\right). \tag{10}$$

However, $q \oplus \vec{w} = p^*$ such that $[p^*] = [p]$, where $[p]$ is the equivalence class of measures up to scale. In other words, p^* is p once it is normalized (2153).

For the normal distribution,

$$\theta^1 = -\frac{1}{2\sigma^2}, \qquad \theta^2 = \frac{\mu}{\sigma^2}, \qquad K(\theta) = \frac{1}{2}\ln\left(-\frac{\pi}{\theta^1}\right) - \frac{(\theta^2)^2}{4\theta^1} \tag{11}$$

On (2155f) I wonder if probability distributions/densities generally wouldn't be better expressed by equivalence classes of measures rather than normalized measures. The following appear to be privileged once you use the exponential parameters rather than simply the probabilities of the categories:

- sufficient statistics
- conjugate priors
- maximum entropy derivation (see Wikipedia)

The log-likelihood of $\mathcal{N}(0,1)$ is 0 because

$$q = q \oplus 0 = e^0 q \tag{12}$$

The log-likelihood of $\mathcal{N}(7,9)$ is

$$\ln \frac{1}{3} + \frac{1}{18}(8x^2 + 14x - 49) \tag{13}$$

Murray and Rice appear to get this wrong, assuming that the coordinates of the origin q are $(0,0)$ so that the log-likelihood of p would be $\theta_p x - K(\theta_p)$, when really it is $\alpha_p x - K(\theta_p)$ with $\theta_q + \alpha_p = \theta_p$ (2160).

Murray and Rice make a momentous claim on page 16:

> Hence P is an exponential family if and only if $\tilde{P}$ is an affine subspace of $\mathcal{M}$.

I could not track with this claim and started reading Jeff Lee's book in order to understand this better (beginning of August, 2017; notes for Murray and Rice end on (2171) and begin for Lee on (2172)).

3 Jeffrey M. Lee: Manifolds and Differential Geometry

3.1 The Tangent Structure (Chapter 2)

3.1.1 A Multivariable Calculus Exercise

Here is a nice exercise for my multivariable calculus students. Consider the vector

$$\vec{OP} = (0.2, 0.3, \sqrt{0.87})^{\intercal} \tag{14}$$

P is on the unit sphere. Provide the equation of the tangent plane. There are two ways to do this (2179):

- use partial derivatives m_x, m_y and the plane equation $z = z_0 + m_x(x - x_0) + m_y(y - y_0)$
- use the dot product and the fact that $\vec{OP}$ is orthogonal to the tangent plane

3.1.2 Lemma 2.4

Showing $(f \circ \gamma_v)'(0) = (f \circ c)'(0)$ on page 59 was a bit of a challenge. Interestingly, understanding this is important for solving exercise 2.16 on page 67 later on. On (2178) I am still carrying on with the assumption on Lee, page 56, that this is the special case of a submanifold of $\mathbb{R}^N$, but given Lee's talk about the "general setting" on the bottom of page 57, this must be incorrect. Show that

$$\left.\frac{d}{dt}\right|_{t=0} (f \circ x_\alpha^{-1} \circ h)(t) = D(f \circ x_\alpha^{-1})(x_\alpha(p)) \cdot \vec{v} \tag{15}$$

with $x_\alpha(p) = \vec{q} \in \mathbb{R}^n$, $p \in \mathcal{M}$, and $h(t) = x_\alpha(p) + t \cdot \vec{v}$, $\vec{v} \in \mathbb{R}^n$. f is a smooth real-valued function defined on $\mathcal{M}$, which has a chart (U_α, x_α). Use the chain rule,

$$D(f \circ x_\alpha^{-1} \circ h)(0) = D(f \circ x_\alpha^{-1})(h(0))D(h)(0) = D(f \circ x_\alpha^{-1})(x_\alpha(p)) \cdot \vec{v} \quad (16)$$

and also

$$(f \circ c)'(0) = (f \circ x_\alpha^{-1} \circ x_\alpha \circ c)'(0) = D(f \circ x_\alpha^{-1})(x_\alpha(p))(x_\alpha \circ c)'(0) \quad (17)$$

(see (2207)).

3.1.3 The Leibniz Law

For the use of the Leibniz Law on page 61 note that $C^\infty(\mathcal{M})$ is the set of real-valued, smooth functions on $\mathcal{M}$. Think of tangent spaces as a straightening of the manifold at a point. $T_p\mathcal{M}$ (kin) is the straightforward geometric interpretation, $T_p\mathcal{M}$ (phys) uses transformation laws, and $T_p\mathcal{M}$ (alg) uses linear functions which fulfill the Leibniz Law. Either way it's conversion therapy for manifolds: straighten them out at a point p by using the linear functions provided by differentiation.

3.1.4 Theorem 2.10

There is a proof on page 63 in Lee that

$$\left(\left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right) \quad (18)$$

is a basis for $T_p\mathcal{M}$ (alg). First, note that g needs to be defined all along u from 0 to u, thus the remark about the convex set. Next, let's have a look at the claim that

$$g = g(0) + \sum g_i u^i \tag{19}$$

based on the FTC. Let $g : x(U) \to \mathbb{R}$ be smooth. Let $h : \mathbb{R} \to x(U)$ be defined by $h(t) = t\vec{u}$. Then

$$h'(t) = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}^{\mathsf{T}} \tag{20}$$

and

$$(g \circ h)'(t) = g'(h(t)) \cdot h'(t) \tag{21}$$

using matrix multiplication. Thus

$$(g \circ h)'(t) = \sum_{i=1}^{n} \frac{\partial g}{\partial u^i}(h(t))u^i \tag{22}$$

and by the FTC

$$\int_0^1 (g \circ h)'(t)\, dt = g(u) - g(0). \tag{23}$$

It follows that

$$\int_0^1 (g \circ h)'(t)dt = \int_0^1 \sum_{i=1}^{n} \frac{\partial g}{\partial u^i}(h(t))u^i\, dt =$$

$$\sum_{i=1}^{n} \left(\int_0^1 \frac{\partial g}{\partial u^i}(t\vec{u})\, dt \right) u^i = \sum_{i=1}^{n} g_i(u)u^i = g(u) - g(0). \tag{24}$$

3.1.5 Exercise 2.16

Show that $T_pf(\beta w_p) = \beta T_pf(w_p)$. Let $[c] = w_p$. Let (U_α, x_α) be an arbitrary chart of $\mathcal{M}$ and $(\tilde{U}_\alpha, \tilde{x}_\alpha)$ an arbitrary chart of $\mathcal{N}$. By definition (see Lee, page 58f),

$$T_pf(\beta w_p) = [f \circ \gamma_1] \text{ and } \beta T_pf(w_p) = [\gamma_2] \tag{25}$$

with $\gamma_1(t) = (x_\alpha^{-1} \circ \phi_1)(t)$ and $\gamma_2(t) = (\tilde{x}_\alpha^{-1} \circ \phi_2)(t)$.

$$\phi_1(t) = x_\alpha(p) + t\beta(x_\alpha \circ c)'(0) \tag{26}$$

$$\phi_2(t) = \tilde{x}_\alpha(q) + t\beta(\tilde{x}_\alpha \circ f \circ c)'(0) \tag{27}$$

(see box A on (2201) and box B on (2203)). Show that for all $g \in C^\infty(\mathcal{N})$

$$(g \circ f \circ \gamma_1)'(0) = (g \circ \gamma_2)'(0) \tag{28}$$

Use the chain rule to separate ϕ_1 and ϕ_2, so the claim reduces to LHS=RHS for

$$\text{LHS} = D(g \circ f \circ x_\alpha^{-1})(x_\alpha(p)) \circ D(\phi_1)(0) \tag{29}$$

$$\text{RHS} = D(g \circ \tilde{x}_\alpha^{-1})(x_\alpha(f(p))) \circ D(\phi_2)(0) \tag{30}$$

with

$$D(\phi_1)(0) = \beta D(x_\alpha \circ c)(0) \tag{31}$$

and

$$D(\phi_2)(0) = \beta D(\tilde{x}_\alpha \circ f \circ c)(0). \tag{32}$$

Now reconstitute (undo the chain rule) for

$$\text{LHS} = \beta D(g \circ f \circ x_\alpha^{-1} \circ x_\alpha \circ c)(0) \tag{33}$$

$$\text{RHS} = \beta D(g \circ \tilde{x}_\alpha^{-1} \circ \tilde{x}_\alpha \circ f \circ c)(0) \tag{34}$$

which are obviously equal to each other (see (2211)). To show that $T_p f(v_p + w_p) = T_p f(v_p) + T_p f(w_p)$ use the same procedure. Let $v_p = [c_v]$ and $w_p = [c_w]$. Then you need to show that

$$(g \circ f \circ \gamma_1)'(0) = (g \circ \gamma_2)'(0) \tag{35}$$

where

$$\gamma_1(t) = x_\alpha^{-1}(x_\alpha(p) + t((x_\alpha \circ c_v)'(0) + (x_\alpha \circ c_w)'(0))) \tag{36}$$

and

$$\gamma_2(t) = \tilde{x}_\alpha^{-1}(\tilde{x}_\alpha(f(p)) + t((\tilde{x}_\alpha \circ f \circ c_v)'(0) + (\tilde{x}_\alpha \circ f \circ c_w)'(0))) \tag{37}$$

Use the distributive law of matrix multiplication to show (35) (see (2213)). Now show that $T_p f$ is a linear isomorphism if f is a diffeomorphism. For surjectivity, let $[\hat{c}] \in T_q\mathcal{N}$. Define $c : t \to (f^{-1} \circ \hat{c})(t)$. Then $[f \circ c] = [\hat{c}]$. For injectivity, let c_1, c_2 be such that $[f \circ c_1] = [f \circ c_2]$ and let $g : \mathcal{M} \to \mathbb{R}$ be smooth. Then $g \circ f^{-1}$ is a smooth function from $\mathcal{N} \to \mathbb{R}$ and therefore

$$(g \circ c_1)'(0) = (g \circ f^{-1} \circ f \circ c_1)'(0) = (g \circ f^{-1} \circ f \circ c_2)'(0) = (g \circ c_2)'(0) \quad (38)$$

which establishes $[c_1] = [c_2]$ (for this proof see (2215)). $\square$

3.1.6 Definition 2.18

On page 68, Lee claims that the definition of $T_p\mathcal{M}(v_p)$ is independent of the representative of v_p. Let $(p, v_1, (U, x_1)), (p, v_2, (U, x_2))$ be two representatives of v_p. Let (V, y) be an arbitrary chart for $\mathcal{N}$ (I suppose we should also show independence of this chart for $\mathcal{N}$). Then $T_p\mathcal{M}(v_p)$ is defined to be represented by $(q, w, (V, y))$ with

$$w = D(y \circ f \circ x^{-1})\big|_{x(p)} \cdot v_p \quad (39)$$

(this is a correction of an omission on Lee's part). The two representatives corresponding to the two charts for $\mathcal{M}$ are $(q, w_1, (V, y))$ and $(q, w_2, (V, y))$ with

$$w_1 = D(y \circ f \circ x_1^{-1}) \cdot v_1 \quad (40)$$

$$w_2 = D(y \circ f \circ x_2^{-1}) \cdot v_2 \quad (41)$$

We need to show that

$$w_2 = D(y \circ y^{-1})\big|_{y(q)} \cdot w_1. \quad (42)$$

You show this by taking $D(y \circ y^{-1} \circ y \circ f \circ x_1^{-1} \circ x_1 \circ x_2^{-1})\big|_{x_2(p)} \cdot v_2$ and using (40) and (41) on the one hand, the chain rule on the other hand (see (2216f)).

3.1.7 Definition 2.20

A word about the definition of differentials on page 69. Using derivations, a tangent vector v_p is a linear map from $C^\infty(\mathcal{M})$ into the real numbers obeying Leibniz's Law. The space of these linear maps is $T_p\mathcal{M}$. Consider a function $f \in C^\infty(\mathcal{M})$. Now define a linear map df assigning to each v_p the real number $v_p \cdot f$. df is a function from $T_p\mathcal{M}$ to the real numbers and depends on f and p.

Consider the dual space (the cotangent space) of $T_p\mathcal{M}$. It is defined as the vector space spanned by $(v_1^*, \ldots, v_n^*)$, where $(v_1, \ldots, v_n)$ is a basis for $T_p\mathcal{M}$ and v_i^* is a function from $T_p\mathcal{M}$ to the real numbers with $v_i^*(v_j) = \delta_{ij}$. Lee's claim is that $df \in T_p^*\mathcal{M}$.

Now look at this in basic calculus. The differential dy is defined as a function assigning to each dx the real number $f'(x)dx$. Think of it as a linear function from the tangent space of f at point a into the real numbers. dx is usually identified with Δx, but now you can see the subtle difference between Δx and dx. If $f : \mathbb{R}^2 \to \mathbb{R}$, then dx is a vector in $\mathbb{R}^2$ and gets mapped to

$$dy = df(p)(dx) = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 \tag{43}$$

where dx_1 is the x-component of dx and dx_2 is the y-component. The z-component of dx is dy by definition (see the diagram on (2223)). For (43), see (2222). The upshot here is that dx is a vector in $T_p\mathcal{M}$, while Δx is a vector in the xy-plane, and we can generalize

$$dy = f'(x)dx \tag{44}$$

to

$$dy = df(p)(dx) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}dx_i \tag{45}$$

where the last expression is only defined for $f : \mathbb{R}^n \to \mathbb{R}$. Now let's see if Lee is correct with his claim that $df \in T_p^*\mathcal{M}$ and that the generalization in (45) holds.

We have a vector space $T_p\mathcal{M}$. Therefore, there is a dual vector space $T_p^*\mathcal{M}$. We know (theorem 2.10) that $T_p\mathcal{M}$ has the basis

$$\left(\frac{\partial}{\partial x^1}\bigg|_p, \ldots, \frac{\partial}{\partial x^n}\bigg|_p \right) \tag{46}$$

The basis for $T_p^*\mathcal{M}$ is $(v_1^*, \ldots, v_n^*)$ such that

$$v_i^* \left(\frac{\partial}{\partial x^j}\bigg|_p \right) = \delta^{ij} \tag{47}$$

If (U, x) is a chart, then the differentials of the coordinate functions $x^1, \ldots, x^n$ fulfill (47) because

$$dx^i\big|_p \left(\frac{\partial}{\partial x^j}\bigg|_p \right) = \frac{\partial x^i}{\partial x^j}(p) = \delta^{ij} \tag{48}$$

From now on, I will follow Lee and write M and N for manifolds, rather than $\mathcal{M}$ and $\mathcal{N}$.

3.1.8 Another Remark on Definition 2.21

Let T_pf be a tangent map from $T_pM \to T_{f(p)}\mathbb{R}$ and $df(p)$ be a differential of f at p. Lee's claim is that T_pf and $df(p)$ basically do the same thing once you identify $T_{f(p)}\mathbb{R}$ with $\mathbb{R}$ using the natural isomorphism on page 66. Therefore, a differential is just a special tangent map into $T_{f(p)}\mathbb{R}$ (identified with $\mathbb{R}$).

To see why this is so note that $T_pf : T_pM \to T_{f(p)}\mathbb{R}$ is the map that sends v_p to $T_pf(v_p)$, a linear map which sends any real-valued function g to the real number

$$(T_p f(v_p))(g) = v_p(g \circ f) \tag{49}$$

Note also that $df(p)$ is a real-valued map sending v_p to $v_p f$. This means we have succeeded once $T_p f(v_p)$ is identified with $v_p f$ via the natural isomorphism $\jmath$. Consider the real number $v_p f$.

$$\jmath(v_p f) = [\hat{c}] \tag{50}$$

where

$$\hat{c}(t) = f(p) + t \cdot (v_p f) \tag{51}$$

Once we show that $[\hat{c}] = T_p f(v_p)$ we are done. Note that $[\hat{c}] \in T_{f(p)}\mathbb{R}$.

Let g be a real-valued function in $C^\infty(M)$ and $[c] = v_p$. Use the interpretation on page 65, the chain rule, and (51) to show that

$$(T_p f(v_p))(g) = v_p(g \circ f) = \left.\frac{d}{dt}\right|_{t=0} (g \circ f \circ c) =$$

$$g'(f(p))f'(p)c'(0) = g'(f(p))(v_p f) = \left.\frac{d}{dt}\right|_{t=0} (g \circ \hat{c}) \tag{52}$$

3.1.9 Definition 2.21

Consider Lee's comment on the bottom of page 69. Let γ be a smooth curve from a real neighbourhood of 0 to $[a, b] = I$ such that

$$[\gamma] = \left.\frac{\partial}{\partial u}\right|_{t_0} \tag{53}$$

According to the interpretation of page 65 (identifying kinematic and algebraic tangent spaces),

$$\left.\frac{d}{dt}\right|_{t=0} f \circ \gamma = \left(\left.\frac{\partial}{\partial u}\right|_{t_0} \right)(f) \tag{54}$$

Let $f = \mathrm{id}$ for

$$\left.\frac{d}{dt}\right|_{t=0} \gamma = \left.\frac{\partial \mathrm{id}}{\partial u}\right|_{t_0} = 1 \tag{55}$$

Now use the same interpretation in reverse for

$$\dot{c}(t_0) \cdot f = \left(T_{t_0} c \left(\left.\frac{\partial}{\partial u}\right|_{t_0} \right) \right) (f) = \left.\frac{d}{dt}\right|_{t=0} f \circ (c \circ \gamma) =$$

$$\left.\frac{d}{dt}\right|_{t=t_0} f \circ c \cdot \left.\frac{d}{dt}\right|_{t=0} \gamma = \left.\frac{d}{dt}\right|_{t=t_0} f \circ c \tag{56}$$

as claimed by Lee. This is a nice example how a mixed use of algebraic and kinematic tangent spaces can be reconciled using the interpretations of page 65. I suppose an alternative way of doing this is to use the algebraic definition of a tangent map $T_p f$ instead (definition 2.19).

3.1.10 Exercise 2.23

$T_p f = 0$ for all $p \in M$ means that for any smooth $g : N \to \mathbb{R}$ and for any $v_p = [c]$ it is true that

$$v_p(g \circ f) = 0 \tag{57}$$

Let p_1 and p_2 be two arbitrary connected points in M such that a curve $c : I \to M$ connects them with $c(a) = p_1, c(0) = p, c(b) = p_2, a < 0 < b$. Then (57) is true for v_p and therefore

$$v_p(g \circ f) = \left.\frac{d}{dt}\right|_{t=0} (g \circ f \circ c) = 0 \tag{58}$$

Define $h(t) = (g \circ f \circ c)(t)$ for an arbitrary g. Claim: $h'(t) = 0$ for all t in the open interval (a, b). Assume that $h'(t_0) \neq 0$. Then define

$$\phi(t) = r_2 t^2 + r_1 t + t_0 \tag{59}$$

with

$$r_1 = \frac{b^2(a - t_0) - a^2(b - t_0)}{ab(b - a)} \tag{60}$$

$$r_2 = \frac{a(b - t_0) - b(a - t_0)}{ab(b - a)}. \tag{61}$$

Then $\phi(a) = a, \phi(0) = t_0, \phi(b) = b$. Also,

$$\phi'(0) = r_1 \neq 0 \tag{62}$$

because $b^2(a - t_0) \neq a^2(b - t_0)$ ($a - t_0$ is the only negative term). Let $\hat{c} = c \circ \phi$ and

$$(g \circ f \circ \hat{c})'(0) = 0 \tag{63}$$

for the same reason as (58). However

$$(g \circ f \circ \hat{c})'(0) = (g \circ f \circ c \circ \phi)'(0) = (g \circ f \circ c)'(\phi(0)) \cdot \phi'(0) \tag{64}$$

and by assumption and (62) both of these factors are not equal to zero. The contradiction gives us the claim that $h'(t) = 0$ for all $t \in (a, b)$. According to the Fundamental Theorem of Calculus, there is a $\hat{s}$ with $a \leq \hat{s} \leq b$ giving us (together with our claim)

$$h(b) - h(a) = h'(\hat{s})(b - a) = 0 \tag{65}$$

Therefore $(g \circ f)(p_1) = (g \circ f)(p_2)$ for arbitrary g. This is only possible if $f(p_1) = f(p_2)$. Therefore, f is locally constant (2229–2237). $\square$

3.1.11 Theorem 2.25

Lee claims that under the assumptions of Theorem 2.25 M and N have the same dimension. Remember that

$$\left(\left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right) \tag{66}$$

is a basis for T_pM. Therefore, $\dim(M) = \dim(T_pM)$ and $\dim(N) = \dim(T_qN)$. If there is a linear isomorphism between two vector spaces V and W, then $\dim(V) = \dim(W)$. For a proof by S.F. Ellermeyer, see

`http://ksuweb.kennesaw.edu/~sellerme/sfehtml/classes/math3260/isomorphicvectorspaces.pdf`

Since T_pf is such a linear isomorphism between T_pM and T_qN, it follows that $\dim(M) = \dim(N)$. $\square$

Notice how Exercise 2.16 and Theorem 2.25 are related. I needed Exercise 2.16 for the claim that T_pf is a *linear* isomorphism. $D(y \circ f \circ x^{-1})(x(p))$ is a linear isomorphism because it is one of the definitions (Definition 2.18) of a tangent map T_pf, and T_pf is a linear isomorphism by assumption. If $y \circ f \circ x^{-1}$ is injective, then f is also injective. Let $\pi_i = x^{-1}(p_i)$ and $f(p_1) = f(p_2)$. Then $\pi_1 = \pi_2$ and therefore $p_1 = p_2$. Restricting the codomain of f to $f(O)$ means that $f|_O$ is a diffeomorphism on O.

3.1.12 Lemma 2.28 Partials Lemma

My work is on (2329–2332). (v, w) is prima facie not a tangent vector. It becomes one by identification with $(T\imath^q + T\imath_p)(v, w)$, see Lee, page 73. $T\imath^q$

is not the inverse of $T_{(p,q)}\mathrm{pr}_1$, as the diagram near the middle of page 73 in Lee suggests. It is the tangent map with respect to the insertion function $\imath^q$. Insertion $\imath^q$ and stripping $T_{(p,q)}\mathrm{pr}_1$ are not inverse to each other. The validity of the lemma is now easily demonstrated.

$$(T_{(p,q)}f(v,w))(g) \overset{(1)}{=} (v,w)(g \circ f) \overset{(2)}{=} v(g \circ f \circ \imath^q) + w(g \circ f \circ \imath_p) \overset{(3)}{=}$$
$$(\partial_1 f_{(p,q)} v)(g) + (\partial_2 f_{(p,q)} w)(g) \tag{67}$$

(1) is true based on the definition of tangent maps (definition 2.19 on page 68) with respect to f. (2) is true based on the identification referred to above and the definition, again, of tangent maps, this time with respect to $\imath^q$ and $\imath_p$. (3) is true based on the definition of partial tangent maps on page 72. □