Three Takes on the Tangent and Cotangent Bundles

Michael Weiss

September 6, 2015

The famous "Blind Men and the Elephant" fable expresses the sentiment of this note. Fully understanding the tangent bundle demands three viewpoints: the formal, the intuitively geometrical, and the classically computational. By formal, I mean set-theoretic, i.e., in the style math has had since we started dwelling in the "Cantorian paradise". By geometrical, I mean visual, or as close as we can come to it: we trade rigor for pictures. By computational, I mean as pre-Cantorian mathematicians might have written it.

1 The Formal Viewpoint

The logical development proceeds by these steps:

1. The notions of *differentiable* and *derivative* are defined for functions $f : \mathbb{R}^n \to \mathbb{R}^m$; more generally, for functions from an open subset $U \subseteq \mathbb{R}^n$ to an open subset $V \subseteq \mathbb{R}^m$. (If you want to get fancy, the

same definitions apply nearly verbatim for functions $f : U \to V$ where U and V are open subsets of Banach spaces; this is how Lang[4] does it. Lang claims that the added generality is worth it, but I'll stick with the more concrete case.) Df (when it exists) is a linear map from R^n to R^m.

2. The space of linear maps $L(\mathbb{R}^n, \mathbb{R}^m)$ is itself a vector space of dimension nm (more concretely, the space of $m \times n$ matrices). So the notions of r-times differentiable and r-th derivative can be defined, and hence the classes $\mathcal{C}^r$ and $\mathcal{C}^\infty$ (smooth) functions.

The second derivative (if it exists) belongs to a space of dimension n^2m, namely $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$, and the r-th derivative (ditto) to one of dimension n^rm. If you prefer, specifying the second derivative at a point requires the n^2m values $\partial^2 y^i/\partial x^j \partial x^k$.

3. The notion of an n-dimensional *differential manifold* M is defined (smooth or $\mathcal{C}^r$ according to preference), via compatible coordinate charts $\phi_U : U \to \phi_U(U) \subseteq \mathbb{R}^n$. The compatibility requirement presupposes the differentiability notions for $\mathbb{R}^n \to \mathbb{R}^n$.

4. The compatibility requirement is tailor-made to support the definition of a *smooth* (or a $\mathcal{C}^r$) map between manifolds, provided the manifold's "niceness" is at least as great as the "niceness" of the maps—e.g., we can define smooth maps between smooth manifolds, but not between $\mathcal{C}^r$ manifolds for $r < \infty$. *Smooth in one chart, smooth in all charts!*

   From now on, for ease of exposition, I'll stick to smooth manifolds.

The legendary heroes after Leibniz and Newton always worked locally. Nonetheless, they did use various charts—polar coordinates, spherical, Euler angles, etc.—and must have converted between them in some way.[1]

[1]Historical footnote: the explicit statement of the chain rule appeared on the late side, it seems not before 1797 in Lagrange according to [5]. I was surprised to read that Euler

Note the bootstrapping technique: we first define *smooth* for functions $\mathbb{R}^n \to \mathbb{R}^m$, and leverage this up to smoothness for functions $M \to N$. The derivative proves more tempermental. Given $f : M \to N$, we can pick charts and look at the composition $\mathbb{R}^n \xrightarrow{\phi^{-1}} M \xrightarrow{f} N \xrightarrow{\psi} \mathbb{R}^m$.[2] But the derivative of $\psi \circ f \circ \phi^{-1}$ depends on the choice of ϕ and ψ, unlike the smoothness. At least the *zeroness* of the derivative at p is chart-independent.

Of course, given $f : M \to N$, the derivative of f can be defined, just not as a map from $\mathbb{R}^n$ to $\mathbb{R}^m$: it needs to take tangent vectors to tangent vectors.

5. The notion of the *tangent vector* at a point $p \in M$ is defined, in one of several equivalent ways (discussed further below):
   - As a *derivation* on the space of *germs* of functions at p.
   - As an equivalence class of smooth curves through p.
   - As n-tuples of numbers "transforming" the right way under coordinate changes.
6. The set of all tangent vectors at p is shown to form a vector space TM_p.
7. The collection of the TM_p is combined into a single object, the tangent bundle TM. The combining employs a seemingly trivial trick: instead of forming the union $\bigcup_{p \in M} TM_p$, we first label each vector with its "tail":

$$TM = \{(p, \mathbf{v}) | p \in M, \mathbf{v} \in TM_p\} = \bigcup_{p \in M} \{p\} \times TM_p$$

never mentions it. I would guess that the free use of differential notation and implicit differentiation served the same purpose.

[2]Really we should restrict attention to open sets $U \subseteq M$, $V \subseteq N$, $\phi(U) \subseteq \mathbb{R}^n$, $\psi(V) \subseteq \mathbb{R}^m$, and look at the composition $\mathbb{R}^n \supseteq \phi(U) \xrightarrow{\phi^{-1}} U \xrightarrow{f} V \xrightarrow{\psi} \psi(V) \subseteq \mathbb{R}^m$. As usual, for greater precision we have to pay a notation tax.

This bit of fussiness matters a lot: it means that a chart on TM naturally includes coordinates for the point p as well as the vector $\mathbf{v}$. So the projection map $\pi : TM \to M$ simply keeps the first n coordinates of an element of TM. (Strictly speaking, you *could* create a funky chart without this property, but why would you want to?)

8. Dual to TM_p we define the space of covectors TM_p^* and the cotangent bundle T^*M, and on top of that build such structures as tensor bundles and bundles of exterior forms.

9. Finally, a *vector field* is defined as a section of TM, and a *diffential form* as a section of T^*M.

Now for the three definitions of *vector*.

Germs. The *germ* of a function $f : U \to \mathbb{R}$ at p (U an open neighborhood of p) is intuitively f restricted to an "infinitesimal neighborhood" of p. Formal definition: consider the family $\mathcal{F}_p$ of all real-valued functions defined on open neighborhoods of p. Define an equivalence relation $\equiv_p$ on $\mathcal{F}_p$ thus: if $f, g \in \mathcal{F}_p$, f defined on U and g on V, then $f \equiv_p g$ iff p has an open neighborhood $W \subseteq U \cap V$ with $f|W = g|W$. A germ at p is an equivalence class under this relation.

If $[f]_p$ is a germ at p, then the value of $[f]_p$ at p is well-defined, being the common value of $g(p)$ for all g in the equivalence class $[f]_p$. Slightly deeper fact: if one $g \in [f]_p$ is infinitely differentiable at p, then *all* $g \in [f]_p$ share this property. So *smooth germ* makes sense. For the germ of a map $f : \mathbb{R}^n \to \mathbb{R}$, the derivative of $[f]_p$ at p makes sense: it's just the common value of $g'(p)$ for all $g \in [f]_p$. The "infinitesimal neighborhood" sobriquet starts to make sense, I hope: $[f]_p$ does not have a well-defined value at any $q \neq p$, but things like derivatives, defined by a limit approaching p, do. (Finally, the notion of germ generalizes to other target spaces without a snag, e.g., "germ of a tensor-valued map".)

Derivations. Let $\mathcal{S}_p$ be the space of smooth germs at p. $\mathcal{S}_p$ is a vector space. A mapping $D : \mathcal{S}_p \to \mathbb{R}$ is a *derivation* at p if it satisfies two properties:

Linearity: $D[af + bg]_p = aD[f]_p + bD[g]_p$ where $a, b \in \mathbb{R}$, $[f]_p, [g]_p \in \mathcal{S}_p$.

Product Rule: $D[fg]_p = [f]_p \, D[g]_p + D[f]_p \, [g]_p$.

It turns out that the set of derivations at p is a vector space of dimension n (assuming p is a point in an n-dimensional manifold M). As Frankel [3] explains, we can think of derivations as directional derivatives.

Writing all those $[\ ]_p$'s can get to be a drag, so some authors (like Frankel) don't bother with germs. They just pick a "small enough" open set U and work inside it. I will follow suit. Indeed, I will be even sloppier, occasionally writing M or $\mathbb{R}^n$ when I should really talk about open subsets. Germs do make things conceptually "cleaner", though. Suppose we have open neighborhoods U and V of p. We can define a tangent vector to be a derivation on the space of smooth maps $U \to \mathbb{R}$, or on the space of smooth maps $V \to \mathbb{R}$. We want these to amount to the same thing. The theory of sheaves (to which germs belong) systematically treats such issues.

Curves. Suppose we have smooth curves $\gamma : (-a, a) \to M$ and $\eta : (-b, b) \to M$ with $\gamma(0) = \eta(0) = p$. We say γ and η are *tangent* at p if composing with a chart ϕ, and thus "transferring the curves to $\mathbb{R}^n$", gives curves in $\mathbb{R}^n$ tangent at $\phi(p)$. In other words, $(\phi \circ \gamma)'(0) = (\phi \circ \eta)'(0)$. Naturally "tangent in one chart, tangent in all charts". Once again we bootstrap: $(\phi \circ \gamma)'(0)$ and $(\phi \circ \eta)'(0)$ are just elements of $\mathbb{R}^n$, so we know what it means for them to be equal.

Using tangency as an equivalence relation, we define a tangent vector to be an equivalence class of smooth curves passing through p (at $t = 0$). We can also define $\gamma'(0)$ to *be* the equivalence class of γ—so $\gamma'(0)$ is a tangent vector. Nice! We can also associate a derivation with $\gamma'(0)$: if $f : U \to \mathbb{R}$ is a smooth function, $p \in U$, then $t \mapsto f(\gamma(t))$ is well-defined for some interval around 0, mapping it to $\mathbb{R}$. We let the "directional derivative" in the direction $\gamma'(0)$ of f (or better, of $[f]_p$) be $(f \circ \gamma)'(0)$. That's a derivation! Building on this one can show that the "curve" definition of tangent vector is equivalent to the "derivation" definition.

***n*-tuples.** No one has done a better job at explaining this approach than Eddington [2]. In a section titled "The mathematical notion of a vector", he writes:

> We have a set of four numbers (A_1, A_2, A_3, A_4) which we associate with some point (x_1, x_2, x_3, x_4) and with a certain system of coordinates. We make a change of the coordinate system, and we ask, What will these numbers become in the new coordinates? The question is meaningless; they do not automatically "become" anything. Unless we interfere with them they stay as they were. But the mathematician may say, "When I am using the coordinates x_1, x_2, x_3, x_4, I want to talk about the numbers A_1, A_2, A_3, A_4; and when I am using x'_1, x'_2, x'_3, x'_4 I find that . . . I shall want to talk about four different numbers A'_1, A'_2, A'_3, A'_4. So for brevity I propose to call both sets of numbers by the same symbol $\mathfrak{A}$."

The mathematician soon realizes he cannot write down the coordinates for $\mathfrak{A}$ in every one of an infinite variety of coordinate systems, and continues:

> "I see that I must alter my plan. I will give you a general rule to find the new values of $\mathfrak{A}$ when you pass from one coordinate system to another. . . "
>
> In mentioning a *rule* the mathematician gives up his arbitrary power. . . He binds himself down to some kind of regularity. Indeed we might have suspected that our orderly-minded friend

> would have some principle in his assignment of different meanings to $\mathfrak{A}$.

Now Eddington asks, can we possibly know what kind of rule the mathematician would adopt, without knowing what problem he is working on? He answers his own question:

> I think we can; it is not necessary to know anything about the nature of his problem...it is sufficient that we know a little about the nature of a mathematician.

Using various assumptions, Eddington narrows the options until only the transformation laws for contravariant and covariant vectors remain. One assumption he states explicitly:

> In technical language the transformations must form a *Group*.

In the following section, Eddington addresses the physical notion of a vector. He introduces the phrase "world-condition" to mean, basically, anything a physicist might want to discuss; he gives force as an example. He warns us:

> A world-condition cannot appear directly in a mathematical equation; only the *measure* of the world-condition can appear.

He next associates a "measure plan" with a coordinate system. Suppose we measure something with plan 1 and call it "world-condition $\mathfrak{A}$", and we measure "the same thing" with plan 2. The measures A_1, A_2, A_3, A_4 and A'_1, A'_2, A'_3, A'_4 that we get *must be related* by one of the mathematician's transformation rules. Otherwise we're not measuring a real physical world-condition, or so Eddington claims. He concludes that tensor analysis is really just souped-up dimensional analysis: if you measure wavelength in meters and in feet (different "measure plans"), the numbers you get better be related by the right conversion factor. With more complicated world-conditions and changes of measure plan, you have to transform whole filing cabinets of numbers, but the basic idea is the same.

Note finally that this third definition for tangent vector involves equivalence classes, as did the second. The equivalence relation is on the set of triples $(U, \phi, \mathbf{v})$ where $\phi : U \to \phi(U) \subseteq \mathbb{R}^n$ is a chart containing p, and $\mathbf{v} \in \mathbb{R}^n$. The tangent vector in TM_p is Eddington's "world-condition"; $\mathbf{v}$ is the n-tuple of measurments of it with the "measure plan" (= chart) ϕ.

2 The Geometrical Viewpoint

In 1955, in the preface to the third edition of his classic work *Die Idee der Riemannschen Fläche* (*The Concept of a Riemann Surface*), Hermann Weyl contrasted the new edition to the first edition of 1913:

> ...I found myself in agreement with the recent trend of topology, to replace the dissection of a manifold by a covering by overlapping neighborhoods. [6]

The early days of topology used triangulations (or more generally dissection into polygons) to construct and understand surfaces. (Look at a proof of Euler's formula $V - E + F = 2$.) We start with a bunch of polygons and glue them together along the edges. So long as we're hand-crafting a topological manifold, no problem: the "patches" can be crumpled anyway (see fig.1). But for a smooth manifold, we use smooth patches, and we don't want to introduce any creases or folds at the seams.

So we use open sets instead of closed, and overlap them. We cut our open patches out of of $\mathbb{R}^2$, or more generally $\mathbb{R}^n$; we picture them as "smooth"

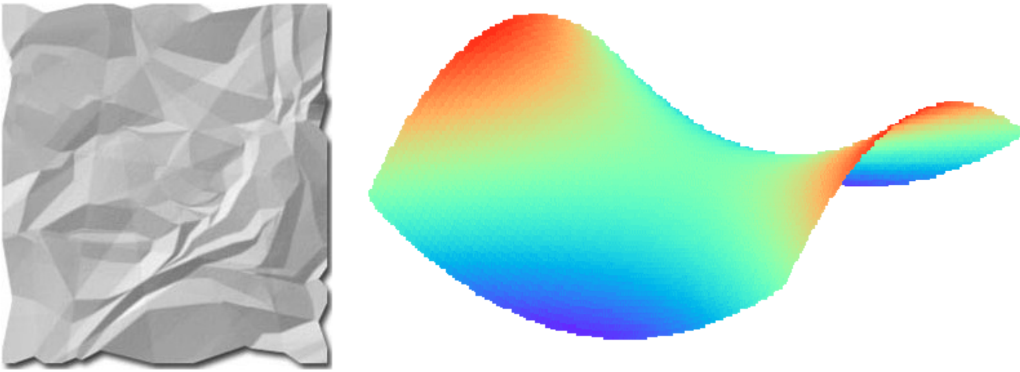

Figure 1: Crumpled and Smooth Patches

patches (again see fig.1), "smoothly" pasted together. What can this mean? If $V \subseteq M$, we imagine we start by cutting the material $\psi(V)$ out of $\mathbb{R}^n$, and attach it to the manifold under construction via the chart map ψ^{-1}. We want $\psi^{-1} : \psi(V) \to V$ to be a "smooth map". But we have as yet no definition for this kind of "smoothness".

Formalization of intuitive concepts often relies on the following strategy. Start by reasoning informally about the intuitive notions, deriving properties capable of formal definition. Postulate these properties for the intuitive concepts. With luck, one ends up with a characterization for which one can prove everything that "ought to be true"; in particular one can formalize the reasoning that led to the postulates in the first place. The three definitions for "tangent vector" all follow this strategy.

This works for smooth manifolds. If all the coordinate charts are "smooth", then each transition function $\psi \circ \phi^{-1}$ should be "smooth" (restricted to $\phi(U \cap V)$, of course). But the transition functions map open subsets of $\mathbb{R}^n$ to open subsets of $\mathbb{R}^n$. We *have* a definition of smoothness for that, so we're in business. We can define "smoothness" for any map between manifolds, and then show that the chart maps are indeed smooth. Thus was born the

At right is a coordinate patch U with chart ϕ. So for $p \in U$ we have $\phi(p) = (x^1, x^2) \in \mathbb{R}^2$ (superscripts); the grid shows the chart lines for x^1 and x^2. The darker lines are level curves of a function f. (The red dot is a saddle-point of f.) Regarding the chart lines as "engraved in the surface", we can think of f as a function of (x^1, x^2), although strictly speaking, $f \circ \phi^{-1}$ is the function of (x^1, x^2).

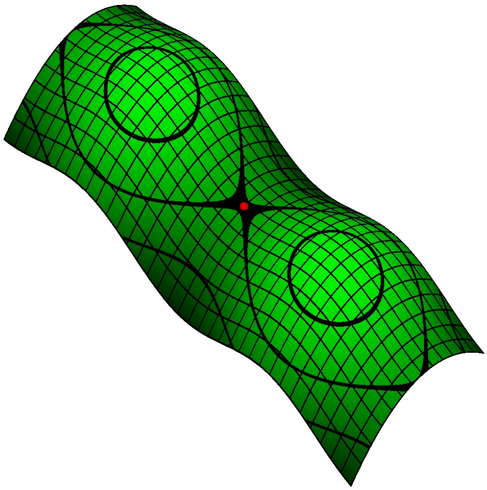

Figure 2: Coordinate Grid

modern definition of differential manifold (perhaps—I haven't researched the history).

The formal definition of manifold emphasizes charts and transition functions, but for visualization, we'd like to get away from that as much as possible. We'd like to picture things *on the manifold* whenever we can.

Charts present no challenge: as Frankel[3] puts it, "*we envisage the coordinates x as being engraved in the manifold M*, just as we see lines of latitude and longitude engraved on our globes." If $f : U \to \mathbb{R}$ is a function on the patch, we can think of it as a function *of the chart coordinates*: see fig.2. Strictly speaking, $f \circ \phi^{-1}$ is that function: $x \mapsto \phi^{-1}(x) \mapsto f(\phi^{-1}(x))$ with route $\phi(U) \to U \to \mathbb{R}$. With a different chart $\psi(p) = y \in \mathbb{R}^n$, we have a different function $f \circ \psi^{-1}$. Classically, we'd write $f(x)$ for the first function and $f(y)$ for the second.

As always with pictures, we have to ask "What's missing from the picture? What's in the picture that's not really there in the math?" Usually, "what's missing" comes from our pathetic confinement to three dimensions. We can visualize surfaces, and $\mathbb{R}^3$—no other 3-manifolds. As for "not really there": pictures of a surface actually show an *embedding* of the surface in 3D. This contains a wealth of deceptive visual information: distances and angles, self-intersections (the Klein bottle), and even smoothness. Fig.1 cheats a bit: the "crumpled patch" has a crumpled embedding, in that the implicit map $U \to \mathbb{R}^3$ isn't smooth. Of course there's a smooth chart $U \to \mathbb{R}^2$—just iron the patch until it lies flat. If we project the patch onto the plane of the paper, that gives us another chart, and the transition function between the two charts isn't smooth.

Tangent vectors. Picture these as little arrows drawn right on the surface. Imagine we are looking at an "infinitesimal patch". Or heed these words from Burke[1], appealing to the "curve" definition:

> For a tangent vector, take that curve in the equivalence class that appears straight in that chart. This special curve in turn can be represented by the piece of it that extends for one unit of parameter change. With an arrowhead on the end, this is the usual idea of a tangent vector, drawn right on the chart of the manifold.

He warns us that the placement of the arrowhead has no intrinsic significance, being chart-dependent.

Cotangent vectors. Picture one of these as a ruling, that is, a bunch of closely spaced short parallel line segments. Think of a tiny piece of a topographical map, blown up so the level curves look straight and parallel. The level curves are marked with "uphill" direction, perpendicular to the lines.

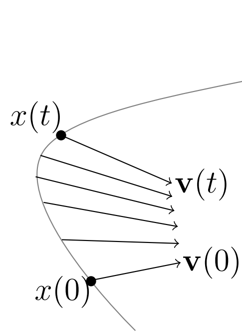

Figure 3: A Curve in TM

Tangent and cotangent bundles TM and T^*M. Of course these are simply the collections of all possible tangent and cotangent vectors, which we can picture, sort of. The dimensionality problem thwarts other approaches: if M is a surface, TM and T^*M have dimension 4, and TTM (for example) has dimension 8. If $\dim M = 1$, we can draw TM and T^*M as surfaces, but 1-dim manifolds tend to be rather humdrum. Nonetheless, one usually gleans some insight from the 1-dim case.

Fig.3 shows the curve $(x(t), \mathbf{v}(t))$ in TM. This is a first step to understanding the so-called double tangent space TTM. Note that $x'(t)$ and $\mathbf{v}(t)$ have no necessary relation to each other, though they both inhabit the same vector space $TM_{x(t)}$. Now, $x'(t)$ tells us *something* about the the vector $(x(t), \mathbf{v}(t))' \in TTM_{(x(t),\mathbf{v}(t))}$; the rest of the story clearly must involve a kind of "acceleration", measuring how $\mathbf{v}(t)$ changes. I will not delve further, as I don't want to discuss connections and covariant differentiation here.

The figure also suggests the natural generalization to a curve in a vector bundle E over M: let the vectors $\mathbf{v}(t)$ point "out of the surface", indicating the fact that now $\mathbf{v}(t) \in E_{x(t)}$, the fiber at $x(t)$, which can be different from $TM_{x(t)}$.

3 The Computational Viewpoint

Back in the days of Euler and friends, computations took place inside a coordinate patch, although you could switch between coordinate systems, using the (multivariable) chain rule. The modern viewpoint usually exacts a "notation tax", which the heroes of old eschewed[3]. I want to elucidate the two perspectives, so we can avoid the tax while still enjoying the fruits of modern conceptual clarity.

Start with the chain rule: say we have $U \xrightarrow{f} V \xrightarrow{g} W$. Initially let's assume U, V, and W are all open subsets of $\mathbb{R}$. The chain rule in modern notation is

$$(g \circ f)' = (g' \circ f)f'$$

but in classical notation

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

which I submit is more handsome and mnenomic. The Leibniz notation revolves around *variables* whereas the modern approach speaks of *functions*. Let's introduce the variables:

$$x \overset{f}{\mapsto} y \overset{g}{\mapsto} z, \quad x \in U, y \in V, z \in W$$

or

$$z = g(y) = g(f(x))$$

The Leibniz expression dz/dx implicitly refers to the function taking us from x to z, likewise for the other two derivatives.

To handle the multivariable case, we just let U, V, W be open sets in $\mathbb{R}^k$, $\mathbb{R}^m$, $\mathbb{R}^n$, and dy/dx is the linear map from the tangent space at $x_0 \in U$ to the tangent space at $y_0 = f(x_0) \in V$, likewise dz/dy and dz/dx.

[3] Gesundheit.

We link up to the material on manifolds by letting M be an open set and let x, y, and z be charts: $x : M \to U$, $y : M \to V$, $z : M \to W$. (We might as well let M be the whole manifold, since an open subset of a manifold is a manifold in its own right.) We imagine a point p "varying" over M, with the chart coordinates "engraved" on M. Thus $x = x(p)$, $y = y(p)$, and $z = z(p)$. We let $p(x)$ stand for the inverse function $x^{-1} : U \to M$, likewise $p(y)$ and $p(z)$. In keeping with fig.2, we can think of y being a function of x, for example, writing the composition $y(p(x))$ as just $y(x)$. This interpretation does lose some generality, since x, y, and z now must be invertible.

Let's look at a classical formula like

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j} \tag{1}$$

We adopt the Einstein summation convention, so there's an implicit $\sum_j$ on the right. Once again, we let x and y be charts on M. So $x = (x^1, \ldots, x^n)$ is a chart on M, and $\partial/\partial x^1, \ldots, \partial/\partial x^n$ is a basis for the tangent space at $p \in M$. Likewise $\partial/\partial y^1, \ldots, \partial/\partial y^n$ is another basis, and our classical formula is just converting between two bases. The same applies for the formula $dy^j = (\partial y^j/\partial x^i)dx^i$.

Another interpretation: suppose we have charts x on M and y on N with $f : M \to N$. The modern version of the classical formula requires a notation tax, since $\partial/\partial y^j$ and $\partial/\partial x^i$ now belong to different tangent spaces. We use the push-forward $f_* : TM_x \to TN_y$, together with the partials of the function $(y \circ f \circ x^{-1}) : (x^1, \ldots, x^n) \mapsto (y^1, \ldots, y^m)$:

$$f_*\left[\frac{\partial}{\partial x^i}\right] = \frac{\partial(y \circ f \circ x^{-1})^j}{\partial x^i}\frac{\partial}{\partial y^j} \tag{2}$$

Likewise, $f^*[dy^j] = (\partial(y \circ f \circ x^{-1})^j/\partial x^i)dx^i$, using the pullback.

Eq.2 arranges the confrontation between tangent vector and basis by pushing forward the tangent vector. Alternately, we could push forward the basis $(\partial/\partial x^i)$. Potential problem: pushing forward a basis doesn't necessarily yield a basis. Let's try it anyway, and see what happens. Start with the classical formula, a variation on (1):

$$\frac{\partial}{\partial y^j} = \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial x^i} \tag{3}$$

The partial derivative $\partial x^i/\partial y^j$ should therefore be the partial of the function $(x \circ f^{-1} \circ y^{-1}) : (y^1, \ldots, y^m) \mapsto (x^1, \ldots, x^n)$. So this interpretation makes sense exactly in the case where f is (locally) invertible (with a differentiable local inverse). So this brings forth essentially nothing new.

References

[1] William L. Burke. *Applied Differential Geometry*. Cambridge University Press, 1985.

[2] Arthur Stanley Eddington. *The Mathematical Theory of Relativity*. Cambridge University Press, 1924.

[3] Theodore Frankel. *The Geometry of Physics*. Cambridge University Press, 3rd edition, 2012.

[4] Serge Lang. *Analysis I & II*. Addison-Wesley, 1968.

[5] Omar Hernández Rodrguez and Jorge M. López Fernández. A Semiotic Reflection on the Didactics of the Chain Rule. *The Montana Mathematics Enthusiast*, 7(2 & 3):321–332, 2010.

[6] Hermann Weyl. *The Concept of a Riemann Surface*. Addison-Wesley, 3rd edition, 1955. Translated by Gerald R. MacLane.