III.C. Linear Functionals I: Derivations and Jacobians

This section offers a *divertissement* on the subject of how linear algebra is applied to multivariable calculus. In one-variable calculus, given a *non*linear function $f\colon \mathbb{R} \to \mathbb{R}$, you look at the tangent line

$$\mathsf{y} = (y - f(a)) = f'(a) \cdot (x - a) = f'(a) \cdot \mathsf{x},$$

which is the graph of the linear transformation approximating f to first order at a. We would like to say something similar for multivariate nonlinear functions $f\colon \mathbb{R}^m \to \mathbb{R}^n$. Rather than just stating it, though, we will try to discover how it arises from geometry. The actual formula for the linear transformation approximating f to first order will come at the end.

Linear functionals. Think back to §I.A, where we considered systems of equations of the form

$$a_1 x_1 + \ldots + a_n x_n = b.$$

How should we think of the left-hand side? Is it $\vec{\alpha} \cdot \vec{x}$? Or a $1 \times n$ matrix (i.e. row) operating on a (column) vector? Or perhaps, a linear transformation

$$\mathbf{a} : \mathbb{R}^n \to \mathbb{R}^1,$$

with the hyperplane (of vectors "perpendicular" to $\vec{\alpha}$)

$$\mathcal{V} = \{a_1 x_1 + \ldots + a_n x_n = 0\} \subseteq \mathbb{R}^n$$

as its kernel (or null space). More generally, any linear transfomation

$$L : V \to \mathbb{R}$$

is called a linear *functional* (or *form*). If V is $\mathbb{R}^n$ then all linear functionals are just "dot product" with a fixed vector $\vec{\alpha} \in \mathbb{R}^n$, as above.

These examples seem pretty boring, huh? The famous mathematician G.H. Hardy called applied mathematics "dehydrated elephants". Seems like the same could be said of abstractified linear algebra. This calls for a strong dose of mathematical culture!

Coordinate systems on manifolds. So let me ask the apparently unrelated question, *How would you differentiate a function on a manifold*, for example on the n-sphere

$$\mathcal{S}^n = \{z_1^2 + \ldots + z_{n+1}^2 = 1\} \subseteq \mathbb{R}^{n+1}\,?$$

First you have to choose a direction against which to look at the rate of change of the function. So it seems plausible that we should denote derivatives by a vector. We will first describe them intrinsically, then explain how to use coordinates to "write them down", exactly as we did with linear transformations.

Let M be an n-dimensional manifold. For our purposes it is enough to think of M as a curvy space that can be patched together from open sets in $\mathbb{R}^n$. Write $C^\infty(M)$ for the smooth functions on M (the ∞ is for "infinitely differentiable"), and take any point $p \in M$. Then a *derivation at* $p \in M$ is any *linear* functional

$$\vec{w} : C^\infty(M) \to \mathbb{R}$$

that obeys the "Leibniz rule"

$$\vec{w}(f \cdot g) \;=\; f(p) \cdot \vec{w}(g) \;+\; g(p) \cdot \vec{w}(f).$$

From this rule you can derive the surprising property that $\vec{w}$ does not assign different values to h_1 and h_2 if they are equal in a small neighborhood of p. For let g be a (smooth) function that is 0 at p, and 1 outside that neighborhood: then $h_1 - h_2 = g \cdot (h_1 - h_2)$, and

$$\vec{w}(h_1 - h_2) = \vec{w}(g \cdot [h_1 - h_2]) =$$

$$= (h_1 - h_2)(p) \cdot \vec{w}(g) + g(p) \cdot \vec{w}(h_1 - h_2) = 0\,;$$

since $\vec{w}$ is linear, $\vec{w}(h_1) = \vec{w}(h_2)$. Also we have

$$\vec{w}(1) = \vec{w}(1 \cdot 1) = 1 \cdot \vec{w}(1) + 1 \cdot \vec{w}(1) = 2 \cdot \vec{w}(1) \quad \Longrightarrow \quad \vec{w}(1) = 0$$

and so $\vec{w}$ kills constants. We denote by $\mathcal{D}_p(M)$ the vector space of derivations at p.

Now let $V_1 \subseteq \mathbb{R}^n$, $U_1 \subseteq M$ be open sets containing the origin and p, respectively, and let

$$\rho_1 : V_1 \longrightarrow U_1$$

be a (smooth, bijective) function taking

$$(0, \ldots, 0) \longmapsto p.$$

You can think of this as a coordinate system (or parametrization map) with origin "at" p. There is a 1-to-1 correspondence between derivations

$$\mathcal{D}_p(M) \longleftrightarrow \mathcal{D}_0(\mathbb{R}^n)$$

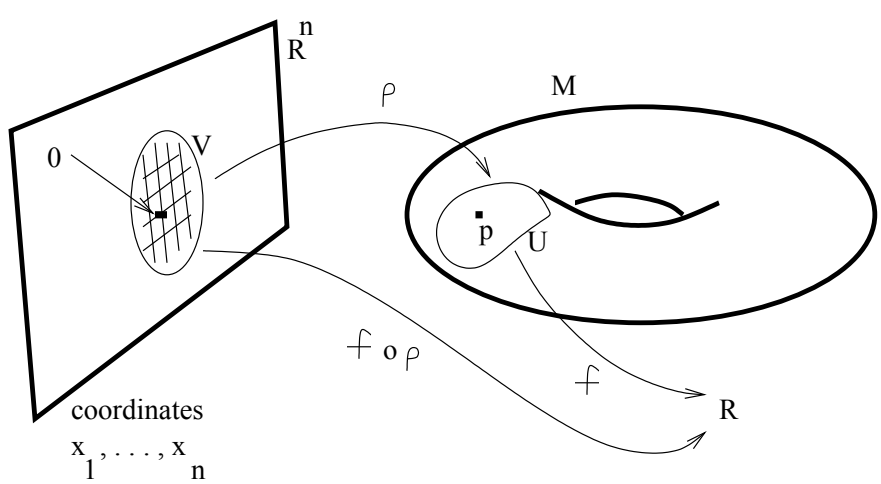
sending

$$\vec{w} \longmapsto (\rho_1^{-1})_* \vec{w}$$
$$(\rho_1)_* \vec{v} \longleftarrow \vec{v},$$

where $(\rho_1)_*$ and $(\rho_1^{-1})_*$ are defined by

$$(\rho_* \vec{v})(f) := \vec{v}(f \circ \rho) \qquad \text{given } f \in C^\infty(M)$$

and $((\rho^{-1})_* \vec{w})(g) = \vec{w}(g \circ \rho^{-1})$. The following picture summarizes the situation:



In words: say I give you a derivation "$\vec{v}$" (at 0) on [functions on] $\mathbb{R}^n$. You can get a derivation "$\rho_* \vec{v}$" (at p) on functions $f \in C^\infty(M)$ by

(i) composing f with ρ to get a function on $\mathbb{R}^n$, then (ii) applying the derivation $\vec{v}$ to the composite $f \circ \rho$. More concisely, you should think of the lower star as "pushing" derivations in the same direction as the map.

Because of this 1-to-1 correspondence, to study derivations $\mathcal{D}_p(M)$ it suffices to understand the possibilities for $\vec{v} \in \mathcal{D}_0(\mathbb{R}^n)$. To this end, let's consider an analytic function on M, so that we may represent it locally by a Taylor series:[5]

$$g(x_1, \ldots, x_n) = a_0 + a_1 x_1 + \ldots + a_n x_n + \sum_{d \geq 2} \left(\sum_{d_1 + \ldots + d_n = d} a_{d_1 \ldots d_n} x_1^{d_1} \ldots x_n^{d_n} \right).$$

On any monomial $x_1^{d_1} \ldots x_n^{d_n}$ of degree ≥ 2, $\vec{v}$ is zero by the Leibniz rule, since we are evaluating[6] at $0 = (0, \ldots, 0)$. So by linearity (and because $\vec{v}$ also kills the constant term),

$$\vec{v}(g) = a_1 \vec{v}(x_1) + \ldots + a_n \vec{v}(x_n),$$

which says that $\vec{v}$ is just the differential operator

$$\vec{v}(x_1) \frac{\partial}{\partial x_1} + \ldots + \vec{v}(x_n) \frac{\partial}{\partial x_n} =: \sum \alpha_i \frac{\partial}{\partial x_i}$$

composed with evaluation at $(0, \ldots, 0)$. So we have

$$\dim(\mathcal{D}_p(M)) = \dim(\mathcal{D}_0(\mathbb{R}^n)) = n,$$

and even a basis
$$\left\{ (\rho_1)_* \frac{\partial}{\partial x_1}, \ldots, (\rho_1)_* \frac{\partial}{\partial x_n} \right\}$$

for $\mathcal{D}_p(M)$ (we sometimes will omit the evaluation at p [or 0] like this). If $\vec{w} \in \mathcal{D}_p(M)$ is any derivation, it can be written uniquely "$\vec{w} = \sum \alpha_i \cdot (\rho_1)_* \frac{\partial}{\partial x_i}$", which means for $f \in C^\infty(M)$

$$\vec{w}(f) \;=\; \sum \alpha_i \frac{\partial (f \circ \rho_1)}{\partial x_i} \Big|_{(0,\ldots,0)}.$$

[5]Not all smooth functions are analytic! See Exercise (1).

[6]For example,

$$\vec{v}(x_1 x_2) = x_1(0) \cdot \vec{v}(x_2) + x_2(0) \cdot \vec{v}(x_1) = 0 \cdot \vec{v}(x_2) + 0 \cdot \vec{v}(x_1) \;=\; 0.$$

There is absolutely nothing special about this basis. let

$$\rho_2 : V_2 \xrightarrow{\sim} U_2$$

be another parametrization, with $\rho_2(0) = p$ for convenience. This gives rise to an alternate basis $\left\{ (\rho_2)_* \frac{\partial}{\partial y_j} \right\}$, with respect to which we can write $\vec{w}$ to get (exactly as above)

$$\vec{w}(f) = \sum \beta_j \frac{\partial (f \circ \rho_2)}{\partial y_j} \Big|_{(0,\ldots,0)}.$$

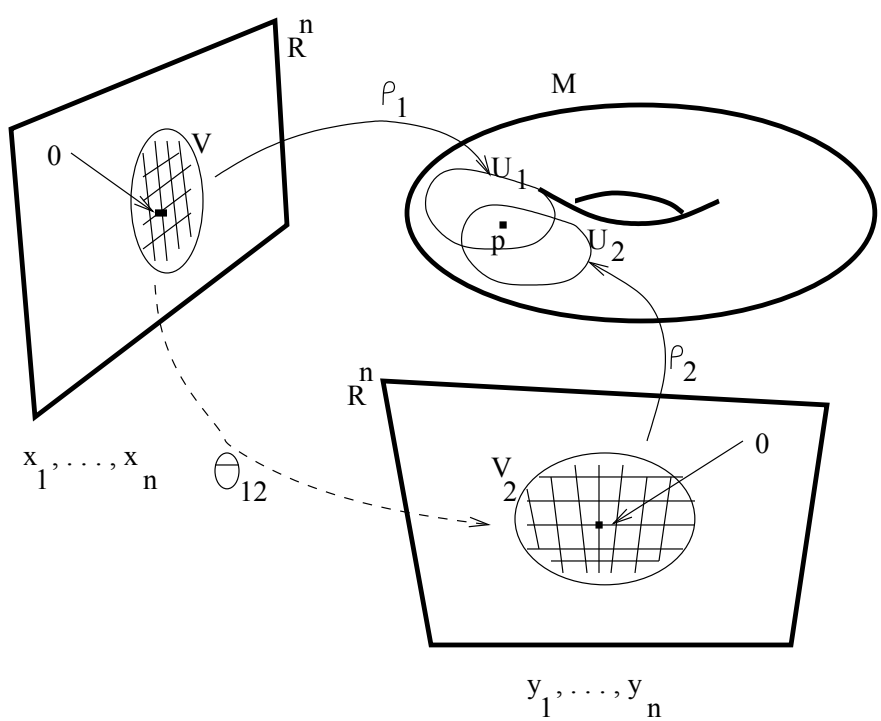
So our goal should be a transition (change of basis) matrix relating the coefficient vectors $\vec{\alpha}$ and $\vec{\beta}$. Restricting if necessary to open subsets of V_1 and V_2 (which map to $U_1 \cap U_2$), we define

$$\theta_{12} := \rho_2^{-1} \circ \rho_1 ,$$

and write

$$\theta_{12}(x_1, \ldots, x_n) =: (y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n))$$

(note: $\theta_{12}(0, \ldots, 0) = (0, \ldots, 0)$). The picture has become:



Now let $g \in C^\infty(V_2)$ be a smooth function, giving us the picture

$$\begin{tikzcd}
 & M & \\
V_1 \arrow[ur, "\rho_1"] \arrow[rr, "\theta_{12}"] \arrow[drr, "g\circ\theta_{12}"'] & & V_2 \arrow[ul, "\rho_2"'] \arrow[d, "g"] \\
 & & \mathbb{R}
\end{tikzcd}$$

Push $\vec{w}$ (on M) to V_1 and V_2 by taking

$$\vec{v}_1 = (\rho_1^{-1})_* \vec{w} = \sum \alpha_i \frac{\partial}{\partial x_i}, \quad \vec{v}_2 = (\rho_2^{-1})_* \vec{w} = \sum \beta_j \frac{\partial}{\partial y_j},$$

and let $\vec{v}_2$ act on g:

$$\begin{aligned}
\vec{v}_2(g) \;&=\; \vec{w}(g \circ \rho_2^{-1}) \;=\; \vec{w}(g \circ \rho_2^{-1} \circ (\rho_1 \circ \rho_1^{-1})) \\
&=\; \vec{w}((g \circ \theta_{12}) \circ \rho_1^{-1}) \;=\; \vec{v}_1(g \circ \theta_{12}).
\end{aligned}$$

That is,

$$\vec{v}_2 = (\theta_{12})_* \vec{v}_1 \;!$$

What this says concretely is:

$$\sum \beta_j \frac{\partial g}{\partial y_j}\Big|_0 \;=\; \vec{v}_2(g) \;=\; \vec{v}_1(g \circ \theta_{12}) \;=\; \sum_i \alpha_i \frac{\partial (g \circ \theta_{12})}{\partial x_i}\Big|_0,$$

which by the Chain rule

$$= \sum_i \alpha_i \sum_j \frac{\partial g}{\partial y_j} \frac{\partial y_j}{\partial x_i}\Big|_0 \;=\; \sum_j \left(\sum_i \alpha_i \frac{\partial y_j}{\partial x_i}\Big|_0 \right) \frac{\partial g}{\partial y_j}\Big|_0.$$

So (switching i and j) we have shown that

$$\beta_i = \sum_j \alpha_j \frac{\partial y_i}{\partial x_j},$$

or (in matrix notation)

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1}\Big|_0 & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_0 \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1}\Big|_0 & \cdots & \frac{\partial y_n}{\partial x_n}\Big|_0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix},$$

and we define $J_{\theta_{12}}|_0$ to be the *Jacobian* matrix $\left[\frac{\partial y_i}{\partial x_j}|_0\right]$. In short,

$$\boxed{\vec{\beta} = J_{\theta_{12}}|_0 \cdot \vec{\alpha}.}$$

Conversely, those derivations $\sum \alpha_i \frac{\partial}{\partial x_i}$ and $\sum \beta_j \frac{\partial}{\partial y_j}$ on distinct coordinate patches which "agree" in the sense of this boxed fomula, correspond to one and the same derivation on M.

In fact, when one studies manifolds, *such equivalence classes of derivations on* $\mathbb{R}^n$ ***are*** (by definition!) *the derivations* $\vec{w}$ *on* M *at* p. How else are you going to differentiate a function at a point on a hyperbolic paraboloid $z = xy$, other than something like "projecting" the function to the xy-plane, and differentiating it there?

What may come as more of a shock is that *this is how you define vectors on* M as well: a tangent vector at p is just a derivation at p! How about smoothly varying vector fields on all of M? Simply cover M with coordinate systems $\rho_\ell : V_\ell \xrightarrow{\sim} U_\ell \subseteq M$, and write down smoothly varying $\sum \alpha_i^\ell(x_1, \ldots, x_n)\frac{\partial}{\partial x_i}$ on each V_ℓ, that "match up" in the sense that $\vec{\alpha}^{\ell_2} = J_{\theta_{\ell_1\ell_2}}\vec{\alpha}^{\ell_1}$.

But suppose I have a submanifold of (say) $\mathbb{R}^3$, like the sphere $\mathcal{S}^2 = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$? Don't we already have a notion of tangent vector at $p \in \mathcal{S}^2$, as vectors in the tangent plane (i.e. a plane through p orthogonal to the gradient of $x^2 + y^2 + z^2 - 1$ at p)? Well, this "extrinsic" approach turns out to be equivalent to the "intrinsic" one we have given, but only works for manifolds embedded in $\mathbb{R}^n$, and is often less convenient even for those.

III.C.1. REMARK. (i) The theory of manifolds takes this philosophy further, defining M abstractly as a topological space with "coordinate chart" maps $\mathbb{R}^n \supset V_\ell \xrightarrow{\rho_\ell} U_\ell \subset M$ which are homeomorphisms (topological isomorphisms), whose images cover M (i.e. $M = \cup U_\ell$), and such that the induced $\theta_{j\ell}$'s on overlaps are C^∞-functions. We don't "assume that ρ_ℓ is C^∞" because that would be circular – the *definition* of a C^∞ function on M is one whose compositions with the ρ_ℓ are all C^∞.

(ii) $\mathcal{D}_p(M)$ is usually written $T_p(M)$, the space of "tangent vectors" = linear differential operators (on M) at p.

Computationally, what have we seen here that is relevant to linear algebra? That one can summarize the first-order (linear) behavior of a nonlinear transformation from $\mathbb{R}^n \to \mathbb{R}^n$ (namely, θ_{12}) by its Jacobian matrix. This is the appropriate way to transform vectors, and it extends to maps $\theta\colon \mathbb{R}^n \to \mathbb{R}^m$, except that J is then an $m \times n$ matrix. Writing $\theta(x_1, \ldots, x_n) = (\theta_1(x_1, \ldots, x_n), \ldots, \theta_m(x_1, \ldots, x_n))$, the induced map θ_* on vectors is linear at each point $p = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$, and has matrix

$$J_\theta\Big|_p = \begin{pmatrix} \frac{\partial\theta_1}{\partial x_1}\Big|_p & \cdots & \frac{\partial\theta_1}{\partial x_n}\Big|_p \\ \vdots & \ddots & \vdots \\ \frac{\partial\theta_m}{\partial x_1}\Big|_p & \cdots & \frac{\partial\theta_m}{\partial x_n}\Big|_p \end{pmatrix}$$

with respect to the "standard basis" (for vectors at each point). From this simpler perspective, the link with differential operators is that they transform (by the chain rule) as vectors, i.e. by the same matrix.

Exercises

(1) For a smooth function G on M, check that any derivation kills $G - L_G$, where $L_G(\underline{x}) = G(0) + G_{x_1}(0)x_1 + \cdots + G_{x_n}(0)x_n$ is the linear approximation (defined only locally), thereby completing the argument above.

(2) Show that the multivariable chain rule for a composition of (nonlinear) maps $F\colon \mathbb{R}^m \to \mathbb{R}^n$ and $G\colon \mathbb{R}^\ell \to \mathbb{R}^m$ can be recast as the statement $J_{F\circ G}|_p = J_F|_{G(p)} \cdot J_G|_p$ (matrix multiplication).

(3) Let $M = \mathcal{S}^2 \subset \mathbb{R}^3$ be given by $x_1^2 + x_2^2 + x_3^2 = 1$, and consider the open sets $U_1 = \mathcal{S}^2 \cap \{x_1 \geq 0\}$ and $U_2 = \mathcal{S}^2 \cap \{x_2 \geq 0\}$. Write down the natural parametrizations ρ_j from the unit disk $D_1 \subset \mathbb{R}^2$ to U_j, then compute θ_{12} and its Jacobian.

(4) Continuing (3), compute the Jacobians of ρ_j, which are 3×2 matrices. Show that (for $P \in D_1$) $J_{\rho_j}|_P$ maps vectors $\vec{v} \in \mathbb{R}^2$ to vectors $(\rho_j)_*\vec{v}$ in the "physical" tangent plane to $\mathcal{S}^2$ at $\rho_j(P)$. How are these Jacobians related to $J_{\theta_{12}}$?