Lecture 6: Computation of the differential

In this lecture we continue to work in the context of standard data (**5.23**). Henceforth we drop the subscripts 'V' and 'W' on the norms, since it is clear from the context which we mean. We also use the operator norm (5.8) on $\mathrm{Hom}(V, W)$ without explicit labeling.

Differentiability and continuity

A differentiable function is continuous, as we now prove.

Theorem 6.1. *Suppose f is differentiable at $p \in U$. Then f is continuous at p.*

Proof. Let $C = \|df_p\|$ be the operator norm of the differential at p. Apply Definition 5.36 with $\epsilon = 1$ to produce $\delta_0 > 0$ such that if $\|\xi\| < \delta_0$, then (5.37) is satisfied. The triangle inequality implies

$$\begin{aligned} \|f(p+\xi) - f(p)\| &\leqslant \|f(p+\xi) - f(p) - df_p(\xi)\| + \|df_p(\xi)\| \\ &< (1+C)\|\xi\|. \end{aligned} \tag{6.2}$$

Given $\epsilon > 0$ choose $\delta = \min\bigl(\delta_0, 1/(1+C)\bigr)$ to satisfy Definition 5.32 of continuity at p. ☐

If the differential of f exists at all points of U, then we can inquire about the continuity of the differential as a map (5.45).

Definition 6.3. If f is differentiable on U and $df\colon U \to \mathrm{Hom}(V, W)$ is continuous, then we say *f is continuously differentiable.*

Functions of one variable

A special case of our general context (**5.23**) is the situation studied in a first analysis course. Then $A = \mathbb{R}$ is the real line and $U \subset \mathbb{R}$ may as well be connected, in which case it is an open interval (a, b) for some real numbers $a < b$. Then $g\colon (a, b) \to B$ is a function of one variable. The simplest situation is $B = \mathbb{R}$, so one function of one variable; if $B = \mathbb{A}^m$, then $g = (g^1, \dots, g^m)$ is m functions of one variable. It is easier in terms of notation to take the codomain B to be an affine space over an arbitrary normed linear space W, and we need this generality later anyhow. Recall (Example 5.25) that we can interpret g as describing a motion in B.

For functions of one variable we define the derivative to be the limit of difference quotients. We foreshadowed the following in (**2.33**).

Definition 6.4. We say *g is old style differentiable at $t_0 \in (a, b)$* if

$$\lim_{h \to 0} \frac{g(t_0 + h) - g(t_0)}{h} \tag{6.5}$$

exists, in which case we notate the limit as $g'(t_0) \in W$.

Multivariable Analysis (M375T, Spring '19), Dan Freed, February 12, 2019

In (6.5) the numerator is the displacement vector between two points of B, and it is scalar multiplied by $1/t$.

Proposition 6.6. *If $g\colon (a,b) \to B$ is old style differentiable at $t_0 \in (a,b)$, then it is differentiable at t_0 and*

$$dg_{t_0}(h) = hg'(t_0), \qquad h \in \mathbb{R}. \tag{6.7}$$

Any linear function $\mathbb{R} \to W$ is determined by its value at 1, which is a vector in W. The statement is that for dg_{t_0} that vector is $g'(t_0)$. We leave the reader to formulate and prove the converse to Proposition 6.6.

Proof. Given $\epsilon > 0$ use the existence of (6.5) to choose $\delta > 0$ such that $(t_0 - \delta, t_0 + \delta) \subset (a,b)$ and if $0 < |h| < \delta$ then

$$\left\| \frac{g(t_0+h) - g(t_0)}{h} - g'(t_0) \right\| < \epsilon. \tag{6.8}$$

Now multiply through by $|h|$ to deduce the estimate in Definition 5.36. (If $h = 0$ that estimate is trivial.) ∎

Computation of the differential

We say a motion $\gamma\colon (a,b) \to A$ has *constant velocity* if it is differentiable and $\gamma'(t)$ is independent of t. In that case γ extends to an affine map $\mathbb{R} \to A$. Given p, ξ there is a unique constant velocity motion $t \mapsto p + t\xi$ with initial position p and velocity ξ.

Now return to our standard data **(5.23)** and fix $p \in U$ and $\xi \in V$. Our task is to compute $df_p(\xi) \in W$, assuming f is differentiable at p. The idea is to use the "tea kettle principle"[1] to reduce to the derivative of a function of one variable, since in that case the differential is computed by the limit of a difference quotient (6.5), and then we have all the techniques and formulas of one-variable calculus available. Let

$$\begin{aligned} \gamma\colon (-r,r) &\longrightarrow U \\ t &\longmapsto p + t\xi \end{aligned} \tag{6.9}$$

be the indicated constant velocity motion, where $r > 0$ is chosen sufficiently small so that the image lies in the open set $U \subset A$.

Theorem 6.10. *If f is differentiable at p, then $f \circ \gamma$ is old style differentiable at 0 and*

$$df_p(\xi) = (f \circ \gamma)'(0). \tag{6.11}$$

Figure 11 depicts the situation in the theorem. In the next lecture we prove a generalization in which γ need not be a constant velocity motion; it need only have initial position p and *initial* velocity ξ.

[1] A mathematician is asked to move a tea kettle from the stove to the sink, which is readily accomplished. The next day the same mathematician is asked to move the tea kettle from the counter to the sink. Solution: move the tea kettle to the stove, thereby reducing the problem to one previously solved.

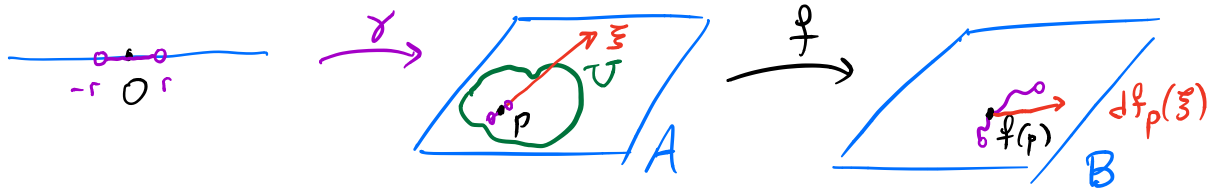

Figure 11. Computing the differential

Proof. We may assume $\xi \neq 0$. Since f is differentiable at p, given $\epsilon > 0$ choose $\delta > 0$ so that if $\eta \in V$ satisfies $\|\eta\| < \delta$, then $p + \eta \in U$ and

$$\|f(p+\eta) - f(p) - df_p(\eta)\| \leqslant \epsilon \frac{\|\eta\|}{\|\xi\|}. \tag{6.12}$$

Then for $0 < |t| < \delta/\|\xi\|$,

$$\left\| \frac{f(p+t\xi) - f(p)}{t} - df_p(\xi) \right\| \leqslant \epsilon. \tag{6.13}$$

This proves the limit of the difference quotient exists and equals $df_p(\xi)$. □

Definition 6.14. We call

$$\left.\frac{d}{dt}\right|_{t=0} f(p + t\xi) \tag{6.15}$$

the *directional derivative of f at p in the direction ξ* and denote it as $\xi f(p)$.

Thus if f is differentiable in U, then given ξ we can differentiate at every point in the direction ξ (using the global parallelism of affine space) to obtain a function

$$\xi f \colon U \to \mathbb{R}. \tag{6.16}$$

Remark 6.17. Theorem 6.10 asserts that if f is differentiable at p, then all directional derivatives at p exist. In the next lecture we prove a converse statement—if directional derivatives exist then f is differentiable—but with restrictions: we assume the domain is finite dimensional and that directional derivatives exist in a neighborhood of p.

Now suppose the domain U is an open subset of the standard affine space $A = \mathbb{A}^n$ for some $n \in \mathbb{Z}^{>0}$. Recall (**2.24**) the standard affine coordinate functions $x^i \colon \mathbb{A}^n \to \mathbb{R}$. In this situation we denote the standard basis elements of the vector space $\mathbb{R}^n$ of translations as

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}. \tag{6.18}$$

The notation is set up so that the directional derivative in the direction of a basis element

$$\frac{\partial}{\partial x^j} f = \frac{\partial f}{\partial x^j} \colon U \to \mathbb{R} \tag{6.19}$$

is the *partial derivative* in the j^{th} coordinate direction. If the codomain $B = \mathbb{A}^m$ is also a standard finite dimensional affine space, then we write $f = (f^1, \dots, f^m)$ for functions $f^i \colon U \to \mathbb{R}$, and then at each $p \in U$ obtain a matrix[2]

$$\left(\frac{\partial f^i}{\partial x^j}(p) \right) \tag{6.20}$$

of partial derivatives. It is the matrix which represents the linear map $df_p \colon \mathbb{R}^n \to \mathbb{R}^m$ in the standard bases.

The operator d and explicit computation

To compute the differential explicitly we observe that the operator d obeys the usual rules of differentiation, as follows from Theorem 6.10 and standard theorems of one-variable calculus. Namely,

(1) d is linear: $d(f_1 + f_2) = df_1 + df_2$
(2) d obeys the Leibniz rule: $d(f_1 \cdot f_2) = df_1 \cdot f_2 + f_1 \cdot df_2$

Notice that we do not exchange the order of the product, which is a good habit since for noncommutative products, as of matrix-valued functions, the same formula applies and one cannot permute factors. Then, after the application of d, we can collect terms and permute factors as allowed. The other basic rule for computing d is the chain rule, which we prove in the next lecture, though of course we already know it for functions of one variable. Using these rules we have a good algorithmic technique and can compute without thinking.

As an example we take $U = A = B = \mathbb{A}^2$, label the standard affine coordinates (r, θ) in the domain and (x, y) in the codomain, and define a function $f \colon \mathbb{A}^2_{(r,\theta)} \to \mathbb{A}^2_{(x,y)}$ by the formulas

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{6.21}$$

We could have written $f(r, \theta) = (r\cos\theta, r\sin\theta)$, but (6.21) is set up for easy computation without thinking, and there are fewer symbols: 'f' does not appear. So simply follow your nose and apply d:

$$\begin{aligned} dx &= dr\ \cos\theta + r\, d(\cos\theta) \\ &= \cos\theta dr - r\sin\theta\, d\theta \end{aligned} \tag{6.22}$$

[2]The superscript j in the denominator is an overall subscript, so i is a superscript and j a subscript. As a matrix i is the row number and j the column number.

The equality $d(\cos\theta) = -\sin\theta\, d\theta$ follows from the chain rule applied to the composition

$$\mathbb{A}^2 \xrightarrow{\theta} \mathbb{R} \xrightarrow{\cos} \mathbb{R}, \tag{6.23}$$

but one gets used to computing without thinking through these justifications. (Do think through them at the beginning!) In the end, applying d to (6.21), we obtain the equations

$$\begin{aligned} dx &= \cos\theta dr - r\sin\theta\, d\theta \\ dy &= \sin\theta dr + r\cos\theta\, d\theta \end{aligned} \tag{6.24}$$

Recall from Remark 5.43 that the differentials $dr, d\theta\colon \mathbb{A}^2 \to (\mathbb{R}^2)^*$ of the affine functions $r, \theta\colon \mathbb{A}^2 \to \mathbb{R}$ are constant on $\mathbb{A}^2$, and they form a basis of $(\mathbb{R}^2)^*$. As in (6.18) the dual basis of $\mathbb{R}^2$ is denoted $\partial/\partial r$, $\partial/\partial\theta$. Evaluate (6.24) on $\partial/\partial r$ to see that the image of the vector $\partial/\partial r$ under the differential of f at (r,θ) is the vector

$$\cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}, \tag{6.25}$$

and the image of the vector $\partial/\partial\theta$ under the differential of f at (r,θ) is the vector

$$-r\sin\theta\frac{\partial}{\partial x} + r\cos\theta\frac{\partial}{\partial y}. \tag{6.26}$$

Remark 6.27. It is worth contemplating this example in some detail to extract some general lessons. We might be tempted to take the image of the (constant) vector *field* $\partial/\partial r$ under df to construct a vector field on $\mathbb{A}^2$. But that is not possible. Observe that $f(0,\theta) = (0,0)$ for all $\theta \in \mathbb{R}$, so to define the value of the supposed image vector field at $(0,0)$ in the codomain we have many choices of which preimage point to use. And (6.25) shows that the vector we obtain is *not* independent of the choice of θ. So there is no well-defined image vector field. If restrict the domain of f to $r > 0$, then each $(x,y) \neq (0,0)$ in the codomain has a collection of preimage points (r,θ) in which any two have the same value of r and values of θ differing by an integer multiple of 2π. Put differently, the preimage is a $\mathbb{Z}$-torsor (Definition 2.22) for the action $n\colon (r,\theta) \to (r, \theta + 2\pi n)$ of $\mathbb{Z}$ on $\mathbb{A}^2_{(r,\theta)}$. Now formula (6.25) shows that the image vector is independent of the choice of preimage, and so there is a well-defined image vector *field*. We depict the image of $\partial/\partial r$ in Figure 12.

Another observation is that the transpose of the differential, which for our general data is a map $df_p^*\colon W^* \to V^*$ or $df^*\colon U \to \mathrm{Hom}(W^*, V^*)$, is what is globally defined always and is what one computes directly. That is one interpretation of (6.24): the right hand side at each (r,θ) is the value of $df^*_{(r,\theta)}$ on dx, dy, respectively.

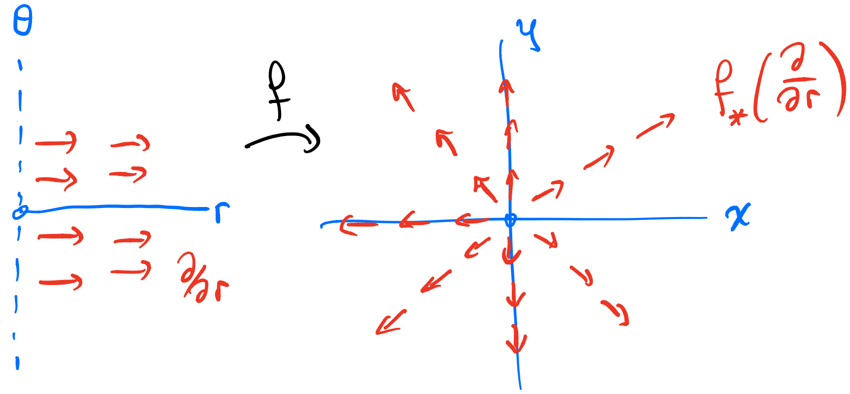

FIGURE 12. Image of the vector field $\partial/\partial r$