Version date: 4/21/17

1 Differentiability

Throughout these notes V and W are finite-dimensional vector spaces (of dimension at least 1) with zero-elements 0_V and 0_W respectively, and with norms $\| \ \|_V$ and $\| \ \|_W$ respectively. The subscripts on the zero-elements and norms will often be dropped when context makes clear which zero-element (that of V, W, or $\mathbf{R}$) or norm is intended.

Definition 1.1 Let $U \subset V$ be an open set, let $F : U \to W$, and let $p \in U$. We say F is *differentiable at* p if F has a *good linear approximation* near p, i.e. if there exists a linear transformation $T : V \to W$ such that

$$\lim_{v \to 0_V} \frac{\|F(p+v) - F(p) - T(v)\|_W}{\|v\|_V} = 0; \tag{1.1}$$

equivalently, if for all $\epsilon > 0$ there exists $\delta > 0$ such that if $v \in V$ and $\|v\| < \delta$,

$$\|F(p+v) - F(p) - T(v)\|_W \leq \epsilon \|v\|_V. \tag{1.2}$$

We say that F *is differentiable* if F is differentiable at every point of U. ▲

Note that (1.1) is equivalent to

$$\lim_{v \to 0_V} \frac{F(p+v) - F(p) - T(v)}{\|v\|_V} = 0_W \ . \tag{1.3}$$

Remark 1.2 The *good linear approximation* referred to in Definition 1.1 is the function $\tilde{F} : q \mapsto F(p) + T(q-p)$. This function is not linear in the sense of "linear transformation", but in the sense of "polynomial of degree at most 1": if we choose bases for V and W, then the component functions of $\tilde{F}$ (relative to the chosen basis of W) are polynomials of degree at most 1 in the coordinate functions of V (relative to the chosen basis of V). This is the only instance in these notes in which "linear function" will mean anything other than what "linear map" or "linear transformation" means in linear algebra. ▲

Claim 1.3 *Let U, F, p be as in Definition 1.1. Then there exists* at most one *linear transformation $T : V \to W$ such that* (1.1) *holds. In particular, if F is differentiable at p then the linear transformation T in equation* (1.1) *is unique.*

Proof: If linear transformations T_1 and T_2 are such that (1.1) holds, then for all $v \in V$

$$(T_2 - T_1)(v) = [F(p+v) - F(p) - T_1(v)] - [F(p+v) - F(p) - T_2(v)] \,.$$

Therefore, for all nonzero $v \in V$,

$$\frac{F(p+v) - F(p) - T_1(v)}{\|v\|} - \frac{F(p+v) - F(p) - T_2(v)}{\|v\|} = \frac{(T_2 - T_1)(v)}{\|v\|} = (T_2 - T_1)\left(\frac{v}{\|v\|}\right),$$

since the linearity of T_1 and T_2 implies that $T_2 - T_1$ is linear as well. Letting $v \to 0$, using the hypothesis that (1.3) is satisfied both with $T = T_1$ and with $T = T_2$, we deduce that

$$0_W = \lim_{v \to 0} (T_2 - T_1) \left(\frac{v}{\|v\|} \right).$$

Hence, by the "Substitution Lemma for limits", for every unit vector $e \in V$ we have

$$0 = \lim_{t \to 0^+} (T_2 - T_1) \left(\frac{te}{\|te\|} \right) = (T_2 - T_1)(e).$$

Since every $v \in V$ is a multiple of some unit vector, linearity implies that $(T_2 - T_1)(v) = 0$ for all $v \in V$, hence that $T_2 = T_1$. ■

Definition 1.4 Let V, W, U, F and p be as above. Suppose F is differentiable at p. The unique linear map $T : V \to W$ satisfying (1.1) is called the *derivative of F at p*, denoted $(DF)_p$ or $DF\big|_p$ in these notes .

Remark 1.5 (Different meaning of "derivative") The derivative of F at p, as defined above, does *not* reduce to the "Calc 1 derivative", i.e. the quantity that (even in Advanced Calc) we have previously called the derivative of F at p in the case $V = W = \mathbf{R}$. The latter is a *number* $F'(p)$, not a linear transformation. However, there is a natural one-to-one correspondence between real numbers and linear transformations $\mathbf{R} \to \mathbf{R}$:

$$\begin{aligned} \mathbf{R} &\longleftrightarrow \mathrm{Hom}(\mathbf{R}, \mathbf{R}), \\ c &\longleftrightarrow \text{the linear map } x \mapsto cx, \end{aligned}$$

i.e. the map "multiplication by c". The derivative of F at p, as defined in Definition 1.4, is the linear map "*multiplication* by $F'(p)$" from $\mathbf{R}$ to $\mathbf{R}$ (see exercise below). Because Definition 1.4 does not reduce to the familiar meaning of "derivative" in the case $V = W = \mathbf{R}$, some authors prefer to call the linear transformation T in (1.1) the *differential* of F at p. ▲

Exercise 1.1 Let $V = W = \mathbf{R}$, and let all other notation be as in Definition 1.1. Show that F is differentiable at p (as defined in Definition 1.1) if and only if $F'(p)$ exists, and that in the differentiable case, the derivative $DF|_p$ is the linear map "multiplication by $F'(p)$" from $\mathbf{R}$ to $\mathbf{R}$.

Earlier this semester, we essentially did the above exercise in class, but did not use the terminology "linear transformation" in defining "good linear approximation", since at that time it would have seemed odd and unmotivated to highlight the fact that $\mathbf{R}$ is a vector space and that the map $x \mapsto F'(p)x$ is a linear transformation from $\mathbf{R}$ to $\mathbf{R}$.

Remark 1.6 Recall that for any fixed $n \in \mathbf{N}$, all norms on $\mathbf{R}^n$ are equivalent. As the student should be able to show, it follows from this that the property of "differentiability at p" and (in the differentiable case) the linear transformation $DF|_p$ are independent of which norms are used on V and W. ▲

Example 1.7 If $F : V \to W$ is a linear map then at each point $p \in U$, $DF\big|_p = F$ ("a linear map is its own derivative [at each point]"). This follows from the uniqueness statement in Claim 1.3 and the fact if F is linear, then for all nonzero $v \in V$ we have

$$\frac{F(p+v) - F(p) - F(v)}{\|v\|} = 0.$$

▲

Remark 1.8 ("Truly advanced" calculus) Most of the "advanced calculus" that we cover in MAA 4211–4212 is actually elementary calculus (Calculus 1-2-3), redone rigorously via an introduction to analysis. One feature of this approach is that it prepares you for more-advanced *analysis*, but there are only a few times in this course—almost all crowded into the end—in which you really get a taste of more-advanced *calculus*. Definitions 1.1 and 1.4 are the doorway to "truly advanced" calculus. these definitions we are *deepening the concept* of derivative. We see, among other things, that linear algebra is inextricably bound to this concept. What is defined in Definition 1.4 is the "grown-up derivative". This deeper concept of *derivative* (at a point) *as a linear transformation* leads to a deeper understanding of various topics in calculus (including the Chain Rule, as we will see later), and paves the way to generalizing calculus on $\mathbf{R}$ and $\mathbf{R}^n$ to calculus on "nonlinear spaces" known as manifolds. Analysis is essential to proving most of the theorems in calculus, but there are *concepts* in calculus that involve only rudimentary analysis (e.g. the notion of of *limit* for functions between normed vector spaces), and whose depth is not in the analytical details. ▲

Exercise 1.2 Let $U \subset V$ be open, let $p \in U$, and let $F_1, F_2, \ldots F_k : U \to W$ be functions that are differentiable at p. Let $F = F_1 + F_2 + \cdots + F_k$. Show that F is differentiable at p, and that $DF|_p = DF_1|_p + DF_2|_p + \cdots + DF_k|_p$.

Definition 1.9 Let $\mathrm{Hom}(V, W)$ denote the space of linear maps $V \to W$ (a vector space of dimension $(\mathrm{dim}V)(\mathrm{dim}W)$). For $T \in \mathrm{Hom}(V, W)$, the *operator norm* of T, denoted $\|T\|_{\mathrm{op}}$, is defined by

$$\|T\|_{\mathrm{op}} = \sup_{\{v \in V:\ \|v\|_V = 1\}} \|T(v)\|_W. \tag{1.4}$$

Note that the value of $\|T\|_{\mathrm{op}}$ may depend on the norms chosen on V and W.

Recall that every (a) linear transformation from one finite-dimensional normed vector space to another is continuous, (b) in any finite-dimensional normed vector space V, the unit sphere $S(V) := \{v \in V : \|v\| = 1\}$ is compact, and (c) any restriction of a continuous function is continuous. Thus, in the setting of (1.4), the function $S(V) \to \mathbf{R}$ defined by $v \mapsto \|T(v)\|$ is continuous (a composition of continuous functions), so the compactness of $S(V)$ implies that this function achieves a maximum value. Thus the supremum in (1.4) is finite (and is actually achieved; "sup" could be replaced by "max").

Exercise 1.3 Show that, as the name and notation suggest, the operator norm is indeed a norm on the vector space $\mathrm{Hom}(V, W)$.

Exercise 1.4 Let $T \in \mathrm{Hom}(V, W)$.

(a) Show that

$$\|T\|_{\mathrm{op}} = \sup_{\{v \in V:\ v \neq 0\}} \left\| T\left(\frac{v}{\|v\|}\right) \right\|.$$

(b) Show that for all $v \in V$,

$$\|T(v)\| \leq \|T\|_{\mathrm{op}}\, \|v\|. \tag{1.5}$$

Proposition 1.10 (Differentiability implies continuity) *Notation as in Definition 1.1. If F is differentiable at p, then F is continuous at p.*

Proof: Assume F is differentiable at p, and let $T = DF|_p$. Let $\delta > 0$ be such that for all $v \in B_\delta(0_V)$ we have $\|F(p+v) - F(p) - T(v)\| \leq \|v\|$; such δ exists by Definitions 1.1 and 1.4. Then for all $q \in B_\delta(p)$, writing $v = q - p$ we have $\|v\| < \delta$, so

$$\begin{aligned}
\|F(q) - F(p)\| = \|F(p+v) - F(p)\| &\leq \|F(p+v) - F(p) - T(v)\| + \|T(v)\| \\
&\leq \|v\| + \|T\|_{\mathrm{op}}\|v\| \\
&= (1 + \|T\|_{\mathrm{op}})\|q - p\|.
\end{aligned}$$

Thus F is Lipschitz at p, hence continuous at p. ■

2 Chain Rule

For the discussion below, it is absolutely critical to remember that "derivative" in these notes has the meaning given in Definition 1.4, *not* the Calculus 1 meaning. Otherwise you will entirely misinterpret the Chain Rule as stated below.

We now add a third finite-dimensional normed vector space $(Z, \| \ \|_Z)$ to the picture so that we can talk about compositions of differentiable functions, and state and prove the Chain Rule Theorem for functions between (subsets of) finite-dimensional vector spaces.

There are several ways of stating the Chain Rule. One way is better than all the others:

The derivative of a composition is the composition of the derivatives.

Some precision is sacrificed in this statement in order to emphasize the elegance and simplicity of the principle. The precise statement is equation (2.1) in the Chain Rule Theorem below.

Theorem 2.1 (Chain Rule Theorem) *Let V, W, Z be finite-dimensional vector spaces. Let $U_1 \subset V, U_2 \subset W$ be open sets, let $F : U_1 \to U_2, G : U_2 \to Z$ be functions, and let $p \in U_1$. Assume that F is differentiable at p and that G is differentiable at $F(p)$. Then the composition $G \circ F$ is differentiable at p, and*

$$D(G \circ F)\big|_p = Dg\big|_{F(p)} \circ DF\big|_p. \tag{2.1}$$

Proof: Let $T = DF\big|_p, q = F(p)$, and $S = DG\big|_q$. Let $\epsilon > 0$. Let $\delta_1, \delta_2 > 0$ be such that for all $w \in W$ and $v \in V$,

$$\text{if } \ \|w\| < \delta_1 \ \text{ then } q + w \in U_2 \text{ and } \ \|G(q + w) - G(q) - S(w)\| \le \epsilon\|w\| \tag{2.2}$$

and

$$\text{if } \|v\| < \delta_2 \text{ then } p + v \in U_1 \text{ and } \ \|F(p + v) - F(p) - T(v)\| \le \min\{\epsilon, 1\}\|v\|; \tag{2.3}$$

such δ_1 and δ_2 exist by Definitions 1.1 and 1.4.

Let $\delta_3 = \min\{\delta_2, \frac{\delta_1}{1+\|T\|_{\rm op}}\}$, and let $v \in V$ be any element with $\|v\| < \delta_3$. Then $\|v\| < \delta_2$, so by (2.3),

$$\|F(p + v) - F(p) - T(v)\| \le \epsilon\|v\| \tag{2.4}$$

and, by the same argument as in the proof of Proposition 1.10,

$$\begin{aligned} \|F(p + v) - F(p)\| &\le (1 + \|T\|_{\rm op})\|v\| & (2.5)\\ &< (1 + \|T\|_{\rm op})\delta_3 \le \delta_1. & (2.6) \end{aligned}$$

Let $w = F(p+v)-F(p)$; thus $F(p+v) = F(p)+w = q+w$. From (2.6) we have $\|w\| < \delta_1$, so, by (2.2) and (2.5),

$$\|G(q + w) - G(q) - S(w)\| \le \epsilon\|w\| \le \epsilon(1 + \|T\|_{\rm op})\|v\|. \tag{2.7}$$

Using (2.7), (2.4), and the linearity of S, we therefore have

$$\begin{aligned} \|(G \circ F)(p + v) - (G \circ F)(p) - (S \circ T)(v))\| &= \|G(q + w) - G(q) - S(T(v))\| & (2.8)\\ &\le \|G(q + w) - G(q) - S(w)\| + \|S(w) - S(T(v))\| &\\ &= \|G(q + w) - G(q) - S(w)\| + \|S(w - T(v))\| &\\ &= \|G(q + w) - G(q) - S(w)\| + \|S\|_{\rm op}\,\|F(p + v) - F(p) - T(v)\| & (2.9)\\ &\le \epsilon(1 + \|T\|_{\rm op})\|v\| + \|S\|_{\rm op}\,\epsilon\|v\| &\\ &= [(1 + \|T\|_{\rm op} + \|S\|_{\rm op})\,\epsilon]\|v\|. & (2.10) \end{aligned}$$

Since ϵ was arbitrary, it follows that $G \circ F$ is differentiable at p, with derivative $S \circ T$. ■

Remark 2.2 The *idea* behind the proof above is the portion from (2.8) through (2.9): starting with an arbitrary v, and writing "$f(p+v) - f(p)$" in place of w, we start with the left-hand side of (2.8) and use the triangle inequality and the linearity of S to attain (2.9). Then we work backwards to see what δ's are needed in order to get from (2.9) to get "(constant $\times\ \epsilon)\|v\|$" in like (2.10).[1] No cleverness is involved; the strategy is completely natural. It is the *more advanced concept of differentiability* that makes this strategy so easy to find; armed with Definitions 1.1 and 1.4, and Definition 1.9 (of which the inequality (1.5) is a simple corollary), this strategy is the first thing we think of. This same natural proof works perfectly well when restricted to the case $V = W = \mathbf{R}$, giving us an alternate proof of our older, single-variable version of the Chain Rule, without resorting to "pulling out of a hat" the cleverly defined function in that earlier proof.

3 Directional derivatives

Definition 3.1 Let U be an open subset of V, let $p \in U$ and let $F : U \to W$ (with no differentiability of F assumed). For $v \in V$, the (generalized) *directional derivative of F at p in direction v* is

$$(D_pF)(v) = \frac{d}{dt}F(p+tv)\bigg|_{t=0} = \lim_{t\to 0} \frac{F(p+tv) - F(p)}{t}$$

if this limit exists. Note that $(D_pF)(v)$, when it exists, is an element of W.

We have inserted "(generalized)" since, unlike in Calculus 3, there is no requirement that v be a unit vector; v can even be the zero vector.

Remark 3.2 Let F and p be given. Trivially, $(D_pF)(0_V) = 0_W$, but the limit in Definition 3.1 may or may not exist for a given nonzero v, and may exist for some nonzero v's but not others. However, if v is a vector for which $(D_pF)(v)$ exists, then, as we are about to show, $(D_pF)(w)$ exists for all multiples w of v, and we have the following homogeneity property:

$$(D_pF)(\lambda v) = \lambda (D_pF)(v) \quad \text{for all } \lambda \in \mathbf{R}. \tag{3.1}$$

[1]By this point in the course, everyone should understand ϵ-δ definitions—in particular, what the quantifiers are telling us *when written in the correct order and location*—well enough to know that if for all $\epsilon > 0$ there exists $\delta > 0$ such that [some stuff] is $\leq$ [any given constant] $\times\ \epsilon$, then for all $\epsilon > 0$ there exists $\delta > 0$ such that [that stuff] is $\leq \epsilon$. Early in the course we would have gone back and rewritten the proof, replacing the ϵ in (2.2) and (2.3) with $\frac{\epsilon}{1+\|T\|_{\text{op}}+\|S\|_{\text{op}}}$ in order to get exactly $\epsilon\|v\|$ in (2.10).

To establish (3.1), note that (3.1) holds trivially if $\lambda = 0$, while for $\lambda \neq 0$ we have

$$\lim_{t\to 0} \frac{F(p+t\lambda v) - F(p)}{t} = \lambda \lim_{t\to 0} \frac{F(p+\lambda t v) - F(p)}{\lambda t} = \lambda (D_pF)(v), \tag{3.2}$$

using the "Substitution Lemma for limits". ▲

Example 3.3 In Definition 3.1, consider the case $V = \mathbf{R}^n, W = \mathbf{R}$; thus F is a real-valued function on an open set in $\mathbf{R}^n$. Writing elements of $\mathbf{R}^n$ as column vectors, let $\{e_j\}_{j=1}^n$ be the standard basis of $\mathbf{R}^n$ (thus $e_j = (0, \ldots, 1, \ldots, 0)^t$, where the 1 in the j^{th} position is the only nonzero coordinate and where the superscript "t" denotes transpose), and let $\{x_j\}_{j=1}^n$ be the standard coordinate functions on $\mathbf{R}^n$. Let $p = (a_1, \ldots, a_n)^t$. Then $p + te_j = (a_1, \ldots, a_{j-1}, a_j + t, a_{j+1}, \ldots, a_n)^t$.

$$\begin{aligned}
(D_pF)(e_j) = \frac{d}{dt}F(p+te_j)\Big|_{t=0} &= \frac{d}{dt}F((a_1, \ldots, a_{ij1}, a_j + t, a_{j+1}, \ldots, a_n)^t)\Big|_{t=0} \\
&= \frac{d}{du}F((a_1, \ldots, a_{i-1}, u, a_{i+1}, \ldots, a_n)^t)\Big|_{u=a_j} \\
&= \frac{\partial F}{\partial x_j}(p),
\end{aligned}$$

the partial derivative of F with respect to x_j (at p), as defined the usual way. Thus, partial derivatives are special cases of directional derivatives. ▲

Proposition 3.4 *If F is differentiable at p then all directional derivatives of F exist at p and*

$$\underbrace{(D_pF)(v)}_{\substack{\textit{directional derivative} \\ \textit{at } p \textit{ in direction } v}} = \underbrace{DF\big|_p(v)}_{\substack{\textit{derivative of } F \textit{ at } p \\ \textit{evaluated at } v}} .$$

Proof: Suppose F is differentiable at p and let $T = DF\big|_p$. Then if $v \neq 0$,

$$\lim_{t\to 0} \frac{F(p+tv) - F(p)}{t} = \lim_{t\to 0} \left(\left[\|v\| \underbrace{\frac{F(p+tv) - F(p) - T(tv)}{t\|v\|}}_{\to 0 \text{ since } F \text{ is differentiable at } p} \right] + T(v) \right) = T(v)$$

■

Corollary 3.5 *Let $U \subset \mathbf{R}^n$ be open, $p \in U$, let $f : U \to \mathbf{R}$ be a function that is differentiable at p, let $\{x_j\}_{j=1}^n$ be the standard coordinates on $\mathbf{R}^n$, and let $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbf{R}^n$. Then*

$$
\begin{aligned}
Df|_p(v) &= \sum_{j=1}^n \frac{\partial f}{\partial x_j}(p)\, v_j &\quad (3.3)\\
&= \underbrace{\left(\frac{\partial f}{\partial x_1}(p),\ \ldots,\ \frac{\partial f}{\partial x_n}(p)\right)}_{1\times n \text{ matrix}} \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}}_{n\times 1 \text{ matrix}}, &\quad (3.4)
\end{aligned}
$$

where matrix multiplication is used on the right-hand side of (3.4), *and* 1×1 *matrices are identified with real numbers.*

Proof: Let $\{e_j\}_{j=1}^n$ be the standard basis of $\mathbf{R}^n$. Then $v = \sum_{j=1}^n v_j e_j$. Using the linearity of $Df|_p$, Proposition 3.4, and Example 3.3, we therefore have

$$Df|_p(v) = \sum_{j=1}^n v_j \ Df|_p(e_j) = \sum_{j=1}^n v_j \ (D_p f)(e_j) = \sum_{j=1}^n v_j \ \frac{\partial f}{\partial x_j}(p),$$

yielding (3.3). ■

4 The case $V = \mathbf{R}^n, W = \mathbf{R}^m$

We now specialize to the concrete case $V = \mathbf{R}^n, W = \mathbf{R}^m$. For purposes of matrix operations that will arise later, we treat elements of $\mathbf{R}^n$ and $\mathbf{R}^m$ as column vectors. When this is inconvenient typographically, so we will write column vectors as transposes of row vectors. If $(a_1, \ldots, a_n)^t$ is in the domain of a function f defined on a subset of $\mathbf{R}^n$, we write simply $f(a_1, \ldots, a_n)$ rather than $f((a_1, \ldots, a_n)^t)$ or $f\left(\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}\right)$.

Throughout this section, unless stated otherwise, $\{e_i\}_{i=1}^n$ and $\{e_i'\}_{i=1}^m$ denote the standard bases of $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, and $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^m$ denote the standard coordinate functions on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. For $1 \le i \le m$, define $\iota_i : \mathbf{R} \to \mathbf{R}^m$ by

$$\iota_i(s) = s e_i' = (0, \ldots, 0, s, 0, \ldots, 0)^t,$$

where the s is in the i^{th} slot.

Observe that each of the functions x_i, y_i, ι_i defined above is a linear map.

Definition 4.1 Let $U \subset \mathbf{R}^n$ be open, and let $F : U \to \mathbf{R}^m$ be a function. For each $i \in \{1, \ldots n\}$ let $f_i = y_i \circ F : U \to \mathbf{R}$ (the i^{th} *component function of* F [with respect to the standard basis of $\mathbf{R}^m$]). Then, as is easily checked,

$$F = \sum_{i=1}^{n} \iota_i \circ f_i \ . \tag{4.1}$$

We may write (4.1) in the more familiar form

$$F = \sum_{j=1}^{m} f_i \, e_i' = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix},$$

with the understanding that this means $F(p) = \sum_{i=1}^{m} f_i(p)\, e_i'$ for all $p \in U$. At any point p for which all the partial derivatives $\frac{\partial f_i}{\partial x_j}(p)$ exist $(1 \leq i \leq n, 1 \leq j \leq m)$, we define the *Jacobian matrix of* F *at* p to be the matrix whose $(ij)^{\text{th}}$ entry is $\frac{\partial f_i}{\partial x_j}(p)$:

$$J_F(p) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \frac{\partial f_1}{\partial x_2}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \frac{\partial f_2}{\partial x_1}(p) & \frac{\partial f_2}{\partial x_2}(p) & \cdots & \frac{\partial f_2}{\partial x_n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \frac{\partial f_m}{\partial x_2}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{pmatrix}.$$

Example 4.2 Consider the case $n = 1$ in Definition 4.1. If the component functions $f_1, \ldots, f_m$ of F are differentiable at $t \in U$, then the Jacobian matrix of F at t is the simply the $m \times 1$ matrix (= column vector)

$$\begin{pmatrix} f_1'(t) \\ \vdots \\ f_m'(t) \end{pmatrix},$$

which is what we called $F'(t)$ when we discussed differential equations (modulo the choices of vector notation and upper/lower-case letters). ▲

Proposition 4.3 *Let* $U, F, \{f_i\}_{i=1}^m$, *and* p *be as in Definition 4.1. Then* F *is differentiable at* p *if and only if each component function* f_i *is differentiable at* p, $1 \leq i \leq m$. *In the differentiable case,*

$$DF|_p = \sum_{i=1}^{m} \iota_i \circ Df_i|_p \ ; \tag{4.2}$$

equivalently,

$$DF|_p(v) = J_F(p)v \quad \text{for all } v \in V, \tag{4.3}$$

where matrix-multiplication is implicit on the right-hand side of this equation.

Thus, Proposition 4.3 yields the following important relation between derivatives and Jacobians:

> *If $F : U \subset \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $p \in U$, then the derivative of F at p is the linear map $\mathbf{R}^n \to \mathbf{R}^m$ given by multiplication by the Jacobian matrix $J_F(p)$.*

Said another way,

> *If $F : U \subset \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $p \in U$, then the Jacobian matrix $J_F(p)$ is the matrix of the linear transformation $DF|_p : \mathbf{R}^n \to \mathbf{R}^m$ with respect to the standard bases of $\mathbf{R}^n$ and $\mathbf{R}^m$.*

For the case $n = m = 1$, we have already seen this fact in Remark 1.5. In this case, the Jacobian matrix $J_F(p)$ is a 1×1 matrix whose sole entry is $F'(p)$. The linear map $x \mapsto F'(p)x$ is exactly multiplication by this 1×1 matrix. Thus, the "Calc 1" derivative of a function $F : (U \subset \mathbf{R}) \to \mathbf{R}$ at p is the 1×1 Jacobian $J_F(p)$.

Proof of Proposition 4.3: First suppose that f_i is differentiable at p, $1 \le i \le m$. Let $i \in \{1, \ldots, m\}$. Since ι_i is linear, Example 1.7 implies that ι_i is differentiable and that $D\iota_i|_{f_i(p)} = \iota_i$. Hence, by the Chain Rule Theorem, $\iota_i \circ f_i$ is differentiable at p, and

$$D(\iota_i \circ f_i)|_p = D\iota_i|_{f_i(p)} \circ Df_i|_p = \iota_i \circ Df_i|_p. \tag{4.4}$$

Since (4.4) holds for each i, and $F = \sum_i \iota_i \circ f_i$, Exercise 1.2 implies the equality (4.2).

Conversely, suppose that F is differentiable at p, and let $i \in \{1, \ldots, m\}$. Then $f_i = y_i \circ F$. Since y_i is linear, the same Chain Rule argument as above shows that f_i is differentiable at p.

This establishes the "if and only if" statement in the Proposition, as well as the equality (4.2) in the differentiable case. For the equivalence between (4.2) and (4.3) (when $DF|_p$ exists), let $v = (v_1, \ldots, v_n)^t \in \mathbf{R}^n$. Then, by Corollary 3.5 (applied to the component functions f_i) and the definition of the maps ι_i,

$$\begin{aligned} DF|_p(v) &= \sum_{i=1}^{m} \iota_i \circ Df_i|_p(v) \\ &= \begin{pmatrix} \sum_{j=1}^n \frac{\partial f_1}{\partial x_j}(p)\, v_j \\ \vdots \\ \sum_{j=1}^n \frac{\partial f_m}{\partial x_j}(p)\, v_j \end{pmatrix} \\ &= J_F(p)v. \end{aligned}$$

Since v was arbitrary, (4.2) and (4.3) are equivalent. ■

Remark 4.4 [**Optional reading**] An alternative derivation of (4.3) is the following. With all notation as in Proposition 4.3, suppose F is differentiable at p. For $1 \le j \le n$, the directional derivative of F at p in direction e_j is

$$\begin{aligned}
(D_pF)(e_j) &= \lim_{t\to 0} \frac{F(p+te_j) - F(p)}{t} \\
&= \lim_{t\to 0} \left(\frac{f_1(p+te_j) - f_1(p)}{t}, \ldots, \frac{f_m(p+te_j) - f_m(p)}{t} \right)^t \\
&= \left(\frac{\partial f_1}{\partial x_j}(p), \cdots, \frac{\partial f_m}{\partial x_j}(p) \right)^t \\
&= \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_j}(p)\, e_i'.
\end{aligned}$$

Since $DF|_p$ is linear, for a general vector $v = \sum_{j=1}^{n} v_j e_j$ we therefore have

$$\begin{aligned}
Df|_p(v) = \sum_{i=1}^{n} v_j \; Df|_p(e_j) &= \sum_{j=1}^{n} v_j (D_pF)(e_j) \\
&= \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{m} \frac{\partial f_i}{\partial x_j}(p) e_i' \right) \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(p) v_j \right) e_i' \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \frac{\partial f_1}{\partial x_2}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \frac{\partial f_2}{\partial x_1}(p) & \frac{\partial f_2}{\partial x_2}(p) & \cdots & \frac{\partial f_2}{\partial x_n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \frac{\partial f_m}{\partial x_2}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.
\end{aligned}$$

[**End of optional reading.**]▲

With notation and hypotheses as in Theorem 2.1, let us now revisit the Chain Rule for the special case $V = \mathbf{R}^n$, $W = \mathbf{R}^m$, and $Z = \mathbf{R}^k$. From Proposition 4.3, for all

$w \in \mathbf{R}^m$ and $v \in \mathbf{R}^n$ we have

$$\begin{aligned} Dg\big|_{f(p)}(w) &= \underbrace{J_g(f(p))}_{k\times m}\underbrace{w}_{\in\mathbf{R}^m} \in \mathbf{R}^k, \\ Df\big|_p(v) &= \underbrace{J_f(p)}_{m\times n}\underbrace{v}_{\in\mathbf{R}^n} \in \mathbf{R}^m, \\ \text{and} \quad D(g\circ f)\big|_p(v) &= \underbrace{J_{g\circ f}(p)}_{k\times n}\underbrace{v}_{\in\mathbf{R}^n} \in \mathbf{R}^k. \end{aligned}$$

Thus Theorem 2.1 implies

$$\underbrace{J_{g\circ f}(p)}_{k\times n} v = D(g\circ f)\big|_p(v) = Dg\big|_{f(p)}\left(Df\big|_p(v)\right) = \underbrace{J_g(f(p))}_{k\times m}\underbrace{J_f(p)}_{m\times n}\underbrace{v}_{\in\mathbf{R}^n} \in \mathbf{R}^k \quad \text{for all } v \in \mathbf{R}^n.$$

Therefore

$$J_{g\circ f}(p) = J_g(f(p))J_f(p), \tag{4.5}$$

i.e. "the Jacobian of a composition is the product of the Jacobians." This is the *second*-best statement of the Chain Rule.

Exercise 4.1 Check that (4.5) is exactly the chain rule you learned in Calculus 3, simply written in matrix notation.

Exercise 4.2 Let $I \subset \mathbf{R}, U \subset \mathbf{R}^m$ be open, let $q \in U$, and suppose that $F : U \to \mathbf{R}^k$ is differentiable. Let $v \in \mathbf{R}^m$, and suppose that $\gamma : I \to \mathbf{R}^m$ is a differentiable function for which $\gamma(0) = q$ and $\gamma'(0) := J_\gamma(0) = v$. Show that

$$\frac{d}{dt}F(\gamma(t))\big|_{t=0} = J_F(q)v = DF|_q(v) = (D_qF)(v).$$

5 Conditions for differentiability

If F is differentiable at p then, as we have seen,

1. The directional derivatives $(D_pF)(v)$ exist for all directions v.
2. For every v, the equality $DF\big|_p(v) = (D_pF)(v)$ holds. Since $DF\big|_p$ is linear, the map $v \mapsto (D_pF)(v)$ must also be linear.

Thus, these are *necessary* conditions for F to be differentiable at p. As the next two examples show, these conditions are *not* sufficient. In these examples, for notational simplicity we write elements of $\mathbf{R}^2$ as row vectors.

Example 5.1 Let $f : \mathbf{R}^2 \to \mathbf{R}$ be the following function (any nonlinear function that is homogeneous of degree 1 would do):

$$f(x,y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Since $f(tx,ty) = tf(x,y)$ we have

$$\left(D_{(0,0)}f\right)((a,b)) = \lim_{t\to 0} \frac{f(t(a,b)) - f((0,0))}{t} = f((a,b)) = \frac{a^3}{a^2+b^2}.$$

Thus, for every (a,b), the directional derivative of f at $(0,0)$ in the direction (a,b) exists. However, the map $(a,b) \to \left(D_{(0,0)}f\right)((a,b))$ is is not linear, so f is not differentiable at $(0,0)$. ▲

The next example shows that even if $(D_pF)(v)$ exists for all v *and* the map $v \mapsto (D_pF)(v)$ is linear, F need not be differentiable at p.

Example 5.2 Consider the function $f : \mathbf{R}^2 \to \mathbf{R}$ given by

$$f(x,y) = \begin{cases} \frac{xy^3}{x^2+y^4} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

For this f we have both

$$f\Big|_{x\text{-axis}} \equiv 0 \qquad \text{and} \qquad f\Big|_{y\text{-axis}} \equiv 0,$$

so $\frac{\partial f}{\partial x}(0,0) = 0 = \frac{\partial f}{\partial y}(0,0)$. Moreover, for any $(a,b) \neq (0,0)$,

$$\frac{f((0,0)+t(a,b)) - f(0,0)}{t} = \frac{f(ta,tb)}{t} = \frac{1}{t}\cdot\frac{t^4ab^3}{t^2a^2+t^4b^4} = \frac{tab^3}{a^2+t^2b^4} \to 0 \qquad \text{as } t \to 0.$$

Thus, all directional derivatives of f exist at $(0,0)$ and are zero (so, in particular, the map $v \mapsto (D_{(0,0)}f)(v)$ is linear). Therefore, if f were differentiable at $(0,0)$ the derivative of f at $(0,0)$ would be the zero-map $\mathbf{R}^2 \to \mathbf{R}$. By the "Substitution Lemma for limits", it would then follow that if $\gamma : \mathbf{R}\setminus\{0\} \to \mathbf{R}^2 \setminus \{0\}$ is any function for which $\lim_{t\to 0}\gamma(t) = (0,0)$ (a curve approaching the origin as $t \to 0$), we must have $\lim_{t\to 0} \frac{f(\gamma(t))-f(0,0)}{\|\gamma(t)\|} = 0$. In particular this would hold for $\gamma(t) = (t^2,t) = (x(t),y(t))$ (approaching the origin along the parabola $x = y^2$). But for this curve γ, we have

$$\lim_{t\to 0}\frac{f(x(t),y(t)) - f(0,0)}{\|(x(t),y(t))\|} = \lim_{t\to 0}\frac{t^5/(2t^4)}{\sqrt{t^4+t^2}} = \lim_{t\to 0}\frac{1}{2\sqrt{1+t^2}}\frac{t}{|t|},$$

which does not exist. Hence f is not differentiable at $(0,0)$. ▲

In view of the previous two examples, one may ask whether there are *any* simple conditions on the directional derivatives of F that guarantee the existence of the derivative of F at a given point? The answer is yes; one such result, stated only for the case $V = \mathbf{R}^n, W = \mathbf{R}^m$, is Proposition 5.3 below. However, bear in mind that the result gives just a *sufficient* conditions for differentiability at a point, not *necessary* condition. Definition 1.1 *cannot* be simplified.

Proposition 5.3 *Let $U \subset \mathbf{R}^n$ be open, $F = (f_1, \ldots, f_m)^t : U \to \mathbf{R}^m$ a function, and $p \in U$. Let $\{x_i\}_{i=1}^n$ be the standard coordinates on $\mathbf{R}^n$. If each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists on some open neighborhood of p, and is continuous at p, then F is differentiable at p.*

Remark 5.4 The condition in Proposition 5.3 is the first condition (sufficient or necessary) we've seen for "F is differentiability at p" that involves knowing differentiability of something at points *other than* p. The fact that it involves any sort of differentiability at points other than p should serve as a reminder that this condition is unlikely to be *necessary* for differentiability at p. The function $f : \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} x^2 \sin(1/x), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

is an example of a function $\mathbf{R} \to \mathbf{R}$ that is differentiable everywhere, but for which the condition in Proposition 5.3 is *not* met at $x = 0$. ▲

Proof of Proposition 5.3: In view of Proposition 4.3, it suffices to prove Proposition 5.3 for the case $m = 1$. Thus, let $f : U \to \mathbf{R}$ be a function such that for $1 \leq j \leq n$, each of the partial derivatives $\frac{\partial f}{\partial x_j}$ exists on an open neighborhood of p, and is continuous at p.

Let $\{e_j\}_{j=1}^n$ be the standard basis of $\mathbf{R}^n$. Define a linear transformation $T : \mathbf{R}^n \to \mathbf{R}$ by $T(\sum_{j=1}^n v_j e_j) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(p) v_j$. Taking the norm on $V = \mathbf{R}^n$ to be the ℓ^1-norm $\| \ \|_1$, and the norm on $W = \mathbf{R}$ to be the standard norm on $\mathbf{R}$, we will show that for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $v \in \mathbf{R}^n$ with $\|v\|_1 < \delta$, (1.3) is satisfied, and therefore that f is differentiable at p. (As noted in Remark 1.6, the choices of norms on V and W do not affect whether f is differentiable at p, so we are free to choose any norms we find convenient.)

Let $\epsilon > 0$. For each $j \in \{1, \ldots, n\}$ let U_j be an open neighborhood of p such that for all $q \in U_j$, $|\frac{\partial f}{\partial x_j}(q) - \frac{\partial f}{\partial x_j}(p)| < \epsilon$. Let $U' = \bigcap_{1 \leq j \leq n} U_j$. Then U' is a finite intersection of open neighborhoods of p, hence an open neighborhood of p. For $r > 0$ and $q \in \mathbf{R}^n$, let $B_r^\infty(q)$ denote the open ball of radius δ and center q in $(\mathbf{R}^n, d_\infty)$, where d_∞ is the ℓ^∞-metric on $\mathbf{R}^n$. Since all norms on $\mathbf{R}^n$ are equivalent, U' contains $B_r^\infty(p)$ for some $r > 0$, hence contains the closed ball $\overline{B}_r^\infty(p)$ for any $r \in (0, r_1)$. Let $\delta > 0$ be such that $\overline{B}_\delta^\infty(p) \subset U'$. By definition of the metric associated with a norm, we have $\overline{B}_\delta^\infty(p) = \{p + v : v \in \overline{B}_\delta^\infty(0)\}$.

Let $\{x_j\}_{j=1}^n$ be the standard coordinates on $\mathbf{R}^n$, and let $v \in V$. For $1 \le j \le n$ let $p_j = x_j(p), v_j = x_j(v)$. (Thus $v = (v_1, \ldots, v_n)^t$ and $|v_j| < \delta$, $1 \le j \le n$.) For $0 \le k \le n$ define $q^{(k)} = p + \sum_{j=1}^k v_j e_j$, the point whose first k coordinates are those of $p+v$ and whose last $n-k$ coordinates are those of p. Then

$$\begin{aligned} f(p+v) - f(p) &= f(q^{(n)}) - f(q^{(0)}) \\ &= [f(q^{(n)}) - f(q^{(n-1)})] + [f(q^{(n-1)}) - f(q^{(n-2)})] + \cdots + [f(q^{(1)}) - f(q^{(0)})]. \end{aligned} \tag{5.1}$$

(Essentially, (5.1), read from the last bracketed expression to the first, says, "Walk from the 'corner' p of a 'cube' to the opposite 'corner' $p+v$ by walking first along an edge parallel to the 1st coordinate axis, then along an edge parallel to the 2nd coordinate axis, etc.") Let $k \in \{1, \ldots, n\}$. Observe that $q^{(k)} = q^{(k-1)} + v_k e_k$. For $t \in [-\delta, \delta]$, the point $z^{(k)}(t) := q^{(k-1)} + te_k$ lies in $\overline{B}_\delta^{\infty}(p)$, on which $\frac{\partial f}{\partial x_k}$ exists and is continuous. But

$$\frac{d}{dt} f(z^{(k)}(t)) = \frac{\partial f}{\partial x_k}(z^{(k)}(t)),$$

so the function $t \mapsto f(z^{(k)}(t))$ is differentiable on an open interval that contains the closed interval with endpoints 0 and v_k. Hence we may apply the Mean Value Theorem and select c_k between 0 and v_k such that

$$f(q^{(k)}) - f(q^{(k-1)}) = f(z^{(k)}(v_k)) - f(z^{(k)}(0)) = \frac{\partial f}{\partial x_k}(z^{(k)}(c_k))v_k. \tag{5.2}$$

Define $\tilde{q}^{(k)} = z^{(k)}(c_k)$. Note that $\tilde{q}^{(k)}$ lies in the ball $B_\delta^\infty(p) \subset U'$, so $|\frac{\partial f}{\partial x_j}(\tilde{q}^{(k)}) - \frac{\partial f}{\partial x_j}(p)| < \epsilon$.

Writing $\tilde{q}^{(k)} = z^{(k)}(c_k)$ for each k, plugging (5.2) into (5.1), and using the definition of T, we have

$$\begin{aligned} |f(p+v) - f(v) - T(v)| &= \left| \sum_{k=1}^n \frac{\partial f}{\partial x_k}(\tilde{q}^{(k)})\, v_k - \sum_{j=1}^n \frac{\partial f}{\partial x_j}(p)\, v_j \right| \\ &= \left| \sum_{j=1}^n \left(\frac{\partial f}{\partial x_j}(\tilde{q}^{(j)}) - \frac{\partial f}{\partial x_j}(p) \right) v_j \right| \\ &\le \sum_{j=1}^n \left| \frac{\partial f}{\partial x_j}(\tilde{q}^{(j)}) - \frac{\partial f}{\partial x_j}(p) \right| |v_j| \\ &\le \sum_{j=1}^n \epsilon |v_j| \\ &= \epsilon \|v\|_1. \end{aligned}$$

Thus for all $v \in \mathbf{R}^n$ with $\|v\|_1 < \delta$, we have $|f(p+v) - f(v) - T(v)| \le \epsilon \|v\|_1$. Since ϵ was arbitrary, it follows that f is differentiable at p. ■

The remainder of this section is optional reading.

Corollary 5.5 *Let $U \subset \mathbf{R}^n$ be open, $F : U \to \mathbf{R}^m$ a function, and $p \in U$. Suppose there is an open neighborhood of U' of p such that for all $v \in \mathbf{R}^n$, the directional derivative $(D_qF)(v)$ exists for every $q \in U'$, and the map $q \mapsto (D_qF)(v)$ is continuous. Then F is differentiable at p.*

Exercise 5.1 (optional) (a) Prove Corollary 5.5.

(b) Strengthen Corollary 5.5 by showing that $\mathbf{R}^n, \mathbf{R}^m$ can be replaced by arbitrary finite-dimensional vector spaces. I.e., prove the following corollary:

Corollary 5.6 *Let V, W be finite-dimensional vector spaces, $U \subset \mathbf{R}^n$ open, $F : U \to \mathbf{R}^m$ a function, and $p \in U$. Suppose there is an open neighborhood of U' of p such that for all $v \in V$, the directional derivative $(D_qF)(v)$ exists for every $q \in U'$, and the map $q \mapsto (D_qF)(v)$ is continuous. Then F is differentiable at p.*

Remark 5.7 Proposition 5.3 is stronger than Corollary 5.5; the Proposition shows that we can deduce differentiability at p from knowing the continuity at p of just all the first partials, of which there are only finitely many, whereas there are infinitely many directional derivatives. However, when V and W are not explicitly $\mathbf{R}^n$ and $\mathbf{R}^m$, there are no "standard coordinates", so the partials used in the Proposition do not make sense. We can always introduce bases for V and W (equivalently, introduce isomorphisms $V \to \mathbf{R}^{\dim(V)}, W \to \mathbf{R}^{\dim(W)}$). A basis of V determines coordinate-functions, while a basis of W determines component-functions $\{f_i\}$ of the map F, so choices of bases allow us to define partial derivatives of component-functions with respect to coordinates on V. However, there are instances in which it is very easy to compute all directional derivatives, and show that they are continuous; introducing a bases and computing partial derivatives of component functions simply becomes extra work. In these instances, Corollary 5.6 can be much more useful than Proposition 5.3. The exercise below illustrates one such instance.▲

Exercise 5.2 Let $V = W = M_{n\times n}$, the space of $n \times n$ matrices. Define $F : V \to V$ by $F(A) = A^2 := AA$. (For any square matrix A and positive integer k, we define $A^k = AA \ldots A$, the product of k copies of A.) (a) Compute $(D_AF)(B)$ for all $A, B \in M_{n\times n}$. (b) Show that for each $B \in V$, the map $A \mapsto (D_AF)(B)$ is continuous. (Hence F is differentiable.)

6 Continuous differentiability

Definition 6.1 *If $F : (U \subset V) \to W$ is differentiable we say F is* continuously differentiable *(on U), or* C^1 *(on U), if the induced map $DF : U \to \mathrm{Hom}(V, W)$ given by $p \mapsto DF\big|_p$ is continuous.*

An immediate corollary of Proposition 5.3 is the following:

Corollary 6.2 *Let $U \subset \mathbf{R}^n$ be open, let $F = (f_1, \ldots, f_m)^t : U \to \mathbf{R}^m$ a function, and let $\{x_i\}_{i=1}^n$ be the standard coordinates on $\mathbf{R}^n$. Then F is continuously differentiable if and only if each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists throughout U and is continuous on U.*

Proof: First assume that F is continuously differentiable. Then $DF|_p$ exists for every $p \in U$, and the map $U \to \mathrm{Hom}(\mathbf{R}^n, \mathbf{R}^m)$ given by $p \mapsto DF|_p$ is continuous. By an earlier exercise, this implies that each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists throughout U and is continuous.

Conversely, assume that each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists throughout U and is continuous on U. By Proposition 5.3, F is differentiable at every point of U. By the same exercise mentioned above, the assumed continuity of the partials implies that the map $U \to \mathrm{Hom}(\mathbf{R}^n, \mathbf{R}^m)$ given by $p \mapsto DF|_p$ is continuous. Hence f is continuously differentiable. ■

Remark 6.3 Because the conditions in Corollary 6.2 are necessary and sufficient for continuous differentiability (not just-plain differentiability!) of $F : (U \subset \mathbf{R}^n) \to \mathbf{R}^m$, the condition "if each of the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists throughout U and is continuous on U" is often taken as the *definition* of "F is continuously differentiable on U", in place of Definition 6.1. (Definition 6.1 is *conceptually* the best definition of "continuous differentiability", but not the easiest definition to apply in practice.) Note, however, that as stated, this alternate definition applies only for functions from (an open subset of) $\mathbf{R}^n$ to $\mathbf{R}^m$. For more general finite-dimensional vector spaces V and W, we must introduce bases, and the associated coordinate functions, in order to make a similar definition. It is not hard to show that, in this more general situation, the continuous-partial-derivatives condition is independent of the choice of bases. ▲

Exercise 6.1 (optional) In the context of Definition 6.1, let $V = \mathbf{R}^n$ and $W = \mathbf{R}^m$, write F as $(f_1, \ldots, f_m)^t$. Show that the map $p \mapsto DF\big|_p$ is continuous if and only if the map

$$p \mapsto J_F(p)$$

is continuous as a map from U to the space $M_{m\times n}$ of $m \times n$ matrices, which in turn is equivalent to all of the real-valued functions $\frac{\partial f_i}{\partial x_j}$ being continuous. (*Suggestion*: Use the fact that $\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$ is a basis of $M_{m\times n}$, where E_{ij} is the $m \times n$ matrix whose $(i,j)^{\text{th}}$ entry is 1 and all of whose other entries are 0.)