Contents

Chapter 1

Review of Advanced Calculus

1.1 Differentiability

Throughout Section 1.1, V and W be finite-dimensional vector spaces with norms $\|\cdot\|_V$ and $\|\cdot\|_W$ respectively. The subscripts on the norms will be dropped from the notation. Wherever $\|\cdot\|$ appears it should be clear from the context whether $\|\cdot\|_V$ or $\|\cdot\|_W$ is intended.

Definition 1.1.1. 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 there exists a linear transformation $T : V \to W$ such that

$$\lim_{h\to 0} \frac{\|f(p+h) - f(p) - T(h)\|_W}{\|h\|_V} = 0. \tag{1.1}$$

Remark 1.1.2. Recall that for $n \in \mathbb{N}$ fixed, all norms on $\mathbb{R}^n$ are equivalent. That is, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are norms on $\mathbb{R}^n$ there there is a constant $C > 0$ such that

$$\frac{1}{C}\|v\|_2 \le \|v\|_1 \le C\|v\|_2 \qquad \text{for all } v \in \mathbb{R}^n.$$

Thus, the concept of differentiability is independent of the choices of norms $\|\cdot\|_V$ and $\|\cdot\|_W$. A proof of the equivalence of all norms on $\mathbb{R}^n$ can be found in Proposition 2.0.11 of the appendix.

Claim 1.1.3. *With all data as in Definition 1.1.1, the linear transformation T in equation (1.1) is unique.*

Proof. If linear transformations T_1 and T_2 are as in (1.1) then for all $h \in V$

$$(T_2 - T_1)(h) = [f(p+h) - f(p) - T_1(h)] - [f(p+h) - f(p) - T_2(h)].$$

Therefore, for all $0 \neq h \in V$,

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

where linearity was used in the final equality. Letting $h \to 0$ shows that

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

Now, if $e \in V$ is any vector of length $\|e\| = 1$, we may choose $h = te$ for nonzero $t \in \mathbb{R}$ to obtain

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

Since e is an arbitrary unit vector and by linearity of T_1 and T_2 we obtain $T_2 \equiv T_1$. ∎

Definition 1.1.4. 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*.

The following notations will be used for the derivative of f at p

$$(Df)_p, \qquad Df\big|_p, \qquad df\big|_p, \qquad (df)_p.$$

Example 1.1.5. Let U be an open subset of V. If $f : U \to W$ is a linear map then at each point $p \in U$, $Df\big|_p(v) = f(v)$ for all v (i.e if f is linear then f is its own derivative). Indeed, consider the difference quotient

$$\frac{f(p+v) - f(p) - T(v)}{\|v\|}.$$

If f is linear then this quotient may be written

$$\frac{f(v) - T(v)}{\|v\|},$$

so choose $T(v) = f(v)$. △

Definition 1.1.6. 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

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

Remark 1.1.7. The limit in Definition 1.1.6 may or may not exist for a given v. Moreover, the limit may exist for some v's but not other v's.

Note that if $(D_p f)(v)$ exists for a given v then then the map span $v \subset V \to W$ given by $u \mapsto (D_p f)(u)$ is homogeneous. Indeed, for any nonzero scalar λ,

$$(D_p f)(\lambda v) = \lambda \lim_{t \to 0} \frac{f(p+t\lambda v) - f(p)}{\lambda t} = \lambda (D_p f)(v). \tag{1.2}$$

In particular, for $v \in V$ given, by choosing $\lambda = \frac{1}{\|v\|}$ we obtain

$$(D_p f)\left(\frac{v}{\|v\|}\right) = \frac{1}{\|v\|}(D_p f)(v).$$

From Definition 1.1.6 we also have $(D_p f)(\lambda v) = \lambda (D_p f)(v)$ when $\lambda = 0$.

Proposition 1.1.8. *If f is differentiable at p then all directional derivatives of f exist at p and*

$$\underbrace{(D_p f)(v)}_{\substack{\text{directional derivative}\\ \text{at } p \text{ in direction } v}} = \underbrace{Df\big|_p(v)}_{\substack{\text{derivative of } f \text{ at } p\\ \text{evaluated at } v}} .$$

Proof. Suppose f is differentiable at p and let $T = Df\big|_p$. Then if $v \neq 0$

$$\lim_{t\to 0} \frac{f(p+tv)-f(p)}{t} = \lim_{t\to 0} \|v\| \underbrace{\frac{f(p+tv)-f(p)-T(tv)}{t\,\|v\|}}_{\to 0 \text{ since } f \text{ is differentiable at } p} + T(v) = T(v)$$

■

Example 1.1.9. Consider the special case that $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ with coordinates $\{x^i\}_{i=1}^n$ and standard bases $\{e_i\}_{i=1}^n$ and $\{e'_i\}_{i=1}^m$

$$e_i = (0,\cdots,0,1,0,\cdots,0)^t, \quad f = (f_1,\cdots,f_m)^t.$$

Suppose $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $p \in \mathbb{R}^n$. The directional derivative of f at p in direction e_i is

$$\begin{aligned}
(D_p f)(e_i) &= \lim_{t\to 0} \frac{f(p+te_i)-f(p)}{t} \\
&= \frac{\partial f}{\partial x^i}(p) \\
&= \left(\frac{\partial f^1}{\partial x^i}(p), \cdots, \frac{\partial f^m}{\partial x^i}(p) \right)^t \\
&= \sum_{j=1}^m \frac{\partial f^j}{\partial x^i}(p) e'_j.
\end{aligned}$$

For a general direction $v = \sum_{i=1}^n v^i e_i$ we have

$$\begin{aligned}
(D_p f)(v) &= \sum_{i=1}^n v^i (D_p f)(e_i) \\
&= \sum_{i=1}^n v^i \sum_{j=1}^m \frac{\partial f^j}{\partial x^i}(p) e'_j \\
&= \sum_{j=1}^m \left(\sum_{i=1}^n \frac{\partial f^j}{\partial x^i}(p) v^i \right) e'_j \\
&= \begin{bmatrix} \frac{\partial f^1}{\partial x^1}(p) & \frac{\partial f^1}{\partial x^2}(p) & \cdots & \frac{\partial f^1}{\partial x^n}(p) \\ \frac{\partial f^2}{\partial x^1}(p) & \frac{\partial f^2}{\partial x^2}(p) & \cdots & \frac{\partial f^2}{\partial x^n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1}(p) & \frac{\partial f^m}{\partial x^2}(p) & \cdots & \frac{\partial f^m}{\partial x^n}(p) \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^m \end{bmatrix}.
\end{aligned}$$

Letting $J_f(p)$ denote the $m \times n$ matrix whose $(ji)^{\text{th}}$ entry is $\frac{\partial f^j}{\partial x^i}(p)$ (i.e. the matrix that appears on the right-hand side of the above string of equalities), we get

$$(D_p f)(v) = J_f(p) v.$$

This equality should be interpreted as follows:

If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $p \in U$ then the derivative of f at p is the linear map $\mathbb{R}^n \to \mathbb{R}^m$ whose representation relative to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is given by multiplication by the Jacobian $J_f(p)$.

△

Exercise 1.1. If f is differentiable at p then f is continuous at p.

1.2 Continuity of Derivatives

Let us start by saying some words about a useful norm topology on the space of linear maps between a pair of fixed vector spaces.

Definition 1.2.1. If V and W are finite-dimensional normed vector spaces then $\mathrm{Hom}(V,W)$ is the $(\dim V)(\dim W)$-dimensional vector space of linear maps $V \to W$. If $\|\cdot\|_V$ and $\|\cdot\|_W$ are the norms on V and W respectively, then the *operator norm* on $\mathrm{Hom}(V,W)$ is given by

$$\|T\|_{\text{op}} = \sup_{\|v\|=1} \|Tv\| = \sup_{v \neq 0} \left\| T \frac{v}{\|v\|} \right\|.$$

If V and W are finite dimensional vector spaces and $T: V \to W$ is linear then T is continuous. Since $\{v \in V : \|v\| = 1\}$ is compact in V (because $\dim V < \infty$), if T is any linear map $V \to W$, the continuity of T ensures that $\{Tv : \|v\| = 1\}$ is compact (hence closed) in W. In particular, if T is a linear map of finite-dimensional vector spaces then $\|T\|_{\text{op}}$ is attained at some $v \in V$ with $\|v\| = 1$.
For $v \neq 0$,

$$\|Tv\| = \|v\| \left\| T\left(\frac{v}{\|v\|}\right) \right\| \leq \|v\| \, \|T\| \, .$$

If S and T are composable linear transformations then for all v,

$$\|(S \circ T)(v)\| \leq \|S\| \, \|Tv\| \leq \|S\| \, \|T\| \, \|v\| \, .$$

If $v \neq 0$ then applying this estimate with v replaced by $v / \|v\|$ gives

$$\left\| (S \circ T) \frac{v}{\|v\|} \right\| \leq \|S\| \, \|T\| \, .$$

Taking the supremum over all nonzero $v \in V$ gives

$$\|ST\| \leq \|S\| \, \|T\| \, .$$

The above inequality says that the operator norm is *submultiplicative*.

With a crash course in the operator-norm topology on spaces of linear transformations behind us, we can talk about continuity of the derivative of a function. Suppose $f: U \to W$ is differentiable (at all points of U). Then f gives rise to a map $Df: U \to \text{Hom}(V, W)$ by $p \mapsto Df\big|_p$ (the derivative of f at p will also be denoted by $D_p f$; see Proposition 1.1.8).

Definition 1.2.2. If $f: U \to W$ is differentiable (at all points of U) we say f is *continuously differentiable* if the induced map $U \to \text{Hom}(V, W)$ given by $p \mapsto D_p f$ is continuous. In this case we write $f \in C^1(U)$.

In the special case $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, $f \in C^1$ if and only if the map

$$p \mapsto J_f(p) = \left[\frac{\partial f^i}{\partial x^j}(p) \right]$$

is continuous. Note that the operator norm topology is the natural topology on $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$.

1.3 How to tell if f is differentiable at p

If f is differentiable at p then

(i) The directional derivatives $(D_p f)(v)$ exist for all directions v.

(ii) For every v, the equality $Df\big|_p(v) = (D_p f)(v)$ holds. Since $Df\big|_p$ is linear $v \mapsto (D_p f)(v)$ must also be linear.

As the next examples show, these conditions are not sufficient (of course, we have already proven the necessity of these conditions).

Example 1.3.1 (Existence of all directional derivatives at p does not imply differentiability at p)**.** Choose any nonlinear function that is homogeneous of degree 1. For example

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

so $f(\lambda x, \lambda y) = \lambda f(x,y)$ and

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

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

The next example shows that even if the directional derivative at p in direction v exists for all v and if the map $v \mapsto D_p f(v)$ is linear, f need not be differentiable at p.

Example 1.3.2. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ 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 $(a,b) \neq 0$,

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

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 is differentiable at $(0,0)$ the derivative of f at $(0,0)$ must be the zero map $\mathbb{R}^2 \to \mathbb{R}$. On the other hand, approach $(0,0)$ along the path $x = y^2$ in the following quotient:

$$\lim_{(x,y)\to 0;\, x=y^2} \frac{f(x,y) - f(0,0)}{\|(x,y)\|} = \lim_{y\to 0} \frac{1}{2\sqrt{1+|y|^2}} \frac{y}{|y|}.$$

This limit (and hence the derivative of f at $(0,0)$) does not exist. △

Given the previous two examples, one may ask the following question: Are there any conditions on the directional derivatives of f that guarantee the existence of the derivative of f? As it turns out, the answer is "yes". The next proposition will address this.

Proposition 1.3.3. *If, for all $v \in V$ the directional derivatives $(D_p f)(v)$ exist for all $p \in U$ and the map $p \mapsto (D_p f)(v)$ is continuous for each fixed v then $f \in C^1(U)$.*

Lemma 1.3.4. *Under the hypotheses of Proposition 1.3.3, the map $v \mapsto (D_p f)(v)$ is linear for each $p \in U$.*

Proof of Lemma. We have already shown that the map $v \mapsto (D_p f)(v)$ is homogenous (see equation (1.2)), so we only need to show that $(D_p f)(u+v) = (D_p f)(u) + (D_p f)(v)$ for all $u, v \in V$. The strategy is to first show that this holds for all u and v with small norm and then remove the norm-smallness requirement.

Fix $p \in U$ and $\varepsilon > 0$. Let $\hat{u}$ and $\hat{v}$ be unit vectors. Choose $\delta > 0$ as in continuity of both $p \mapsto (D_p f)(\hat{v})$ and $p \mapsto (D_p f)(\hat{u})$. That is if $\|q - p\| < \delta$ then both

$$\left\|(D_p f)(\hat{v}) - (D_q f)(\hat{v})\right\| < \varepsilon \qquad \text{and} \qquad \left\|(D_p f)(\hat{u}) - (D_q f)(\hat{u})\right\| < \varepsilon.$$

If $u = \lambda\hat{u}$ and $v = \alpha\hat{v}$ then by homogeneity of directional derivatives,

$$\left\|(D_p f)(v) - (D_q f)(v)\right\| < \varepsilon \|v\| \quad \text{and} \quad \left\|(D_p f)(u) - (D_q f)(u)\right\| < \varepsilon \|u\|.$$

Therefore, if $u = \lambda\hat{u}$ and $v = \alpha\hat{v}$ with $\|u\| < \delta/2$ and $\|v\| < \delta/2$, then $p + su + tv \in B_\delta(p)$ whenever s and t are real numbers satisfying $|s|, |t| \leq 1$.

Now consider

$$\begin{aligned} f(p+u+v) - f(p) - (D_p f)(u) - (D_p f)(v) &= \left[f(p+u+v) - f(p+u) - (D_p f)(v)\right] \\ &\quad + \left[f(p+u) - f(p) - (D_p f)(u)\right] \end{aligned} \tag{1.3}$$

and $g(t) = f(p+tu)$ for $0 \leq t \leq 1$. We have

$$g'(t_0) = \lim_{h\to 0} \frac{f(p+t_0u+hu) - f(p+t_0u)}{h} = (D_{p+t_0u}f)(u).$$

This implies

$$g'(t) = \underbrace{(D_{p+tu}f)(u)}_{\text{continuous in } t}.$$

Now,

$$f(p+u) - f(p) = g(1) - g(0) = \int_0^1 \frac{d}{dt} g(t)\, dt = \int_0^1 (D_{p+tu}f)(u)\, dt$$

so

$$f(p+u) - f(p) - (D_pf)(u) = \int_0^1 \left[(D_{p+tu}f)(u) - \underbrace{(D_pf)(u)}_{\text{independent of } t} \right] dt$$

and

$$\left\| f(p+u) - f(p) - (D_pf)(u) \right\| \leq \int_0^1 \left\| (D_{p+tu}f)(u) - (D_pf)(u) \right\| dt \leq \varepsilon \left\| u \right\|,$$

the final estimate holding as $p+tu \in B_\delta(p)$.

Next, we handle the term $f(p+u+v) - f(p+u) - (D_pf)(v)$ in equation (1.3). By a similar argument we have

$$\begin{aligned} \left\| f(p+u+v) - f(p+u) - (D_pf)(v) \right\| &= \left\| \int_0^1 \left[(D_{p+u+tv}f)(v) - (D_pf)(v) \right] dt \right\| \\ &\leq \int_0^1 \left\| (D_{p+u+tv}f)(v) - (D_pf)(v) \right\| dt \\ &\leq \varepsilon \left\| v \right\|. \end{aligned}$$

Thus, in view of equation (1.3), by taking norms we have

$$\left\| f(p+u+v) - f(p) - (D_pf)(u) - (D_pf)(v) \right\| \leq \varepsilon(\|u\| + \|v\|).$$

Replacing u by tu and v by tv with $|t| \leq 1$ we have

$$\frac{\left\| f(p+t(u+v)) - f(p) - t\left[(D_pf)(u) + (D_pf)(v) \right] \right\|}{|t|} \leq \frac{\varepsilon\, |t|\, (\|u\| + \|v\|)}{|t|} = \varepsilon(\|u\| + \|v\|).$$

Therefore, as $t \to 0$, the quotient becomes arbitrarily small and we get

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{f(p+t(u+v)) - f(p) - t\left[(D_pf)(u) + (D_pf)(v) \right]}{t} \\ &= \lim_{t\to 0} \underbrace{\frac{f(p+t(u+v)) - f(p)}{t}}_{\to (D_pf)(u+v) \text{ by hypothesis}} - \left[(D_pf)(u) + (D_pf)(v) \right] \end{aligned}$$

Therefore, if u is a scalar multiple of $\hat{u}$ with $\|u\| < \delta/2$ and if v is a scalar multiple of $\hat{v}$ with $\|v\| < \delta/2$ then

$$(D_p f)(u+v) = (D_p f)(u) + (D_p f)(v).$$

Next we remove the restriction on the smallness conditions on $\|u\|$ and $\|v\|$. Given u and v (of possibly large norms) choose a scalar λ such that both $\|u/\lambda\| < \delta/2$ and $\|v/\lambda\| < \delta/2$. Then

$$(D_p f)\left(\frac{u}{\lambda} + \frac{v}{\lambda}\right) = (D_p f)\left(\frac{u}{\lambda}\right) + (D_p f)\left(\frac{v}{\lambda}\right).$$

The equality $(D_p f)(u+v) = (D_p f)(u) + (D_p f)(v)$ now follows from homogeneity of directional derivates. ■

Remark 1.3.5. Suppose instead of assuming existence and continuity of $p \mapsto (D_p f)(v)$ for all v, we only assume for $v \in \{e_1, \cdots, e_n\}$ (a basis for V). Then

$$(D_p f)(u+v) = (D_p f)(u) + (D_p f)(v) \quad \text{for all } u \in \text{span}\{e_1\}, v \in \text{span}\{e_2\}.$$

Therefore, $(D_p f)$ restricted to $\text{span}\{e_1, e_2\}$ is linear. Inductively, suppose $u \in \text{span}\{e_1, e_2\}$, $v \in \text{span}\{e_3\}$. We can show that the restriction of $(D_p f)$ to $\text{span}\{e_1, e_2, e_3\}$ is linear. Repeating this procedure a finite number of times we can obtain the linearity of $(D_p f)$ (the "restriction" of $(D_p f)$ to $\text{span}\{e_1, \cdots, e_n\}$).

Proof of Proposition 1.3.3. Fix $p \in U$. We want to show there exists a linear map $T : V \to W$ such that

$$\frac{\|f(p+v) - f(p) - T(v)\|}{\|v\|} \to 0 \qquad \text{as } \|v\| \to 0.$$

Given that $(D_p f)(v)$ exists for all v and that the map $v \mapsto (D_p f)(v)$ is linear, the obvious choice for T is $T(v) = \sum_{i=1}^n (D_p(f)(v^i e_i)$, where $\{e_i\}_{i=1}^n$ is a basis for V relative to which $v = \sum_{i=1}^n v^i e_i$. We showed in the lemma that given $\varepsilon > 0$ and unit vector $\hat{u}$, there exists $\delta > 0$ such that

$$\left\| f(q+u) - f(q) - (D_q f)(u) \right\| \le \varepsilon \|u\| \tag{1.4}$$

whenever $\|u\| < \delta$ and $u \in \text{span}\{\hat{u}\}$. Let $v = \sum_{i=1}^n v^i e_i$ satisfy $\min(\|v\|_\infty, \|v\|) < \delta$ so that $p + v \in B_\delta(p)$. For ease of notation, we set for $j = 1, \cdots, n$,

$$w_j = \sum_{i=1}^j v^i e_i,$$

the projection of v onto $\text{span}\{e_1, \cdots, e_j\}$ and $w_0 = 0$. Note that for each j, $w_j \in B_\delta(p)$ so we can use estimate (1.4) with q replaced by $p + w_{j-1}$ and u replaced by $v^j e_j$. Accordingly, we have

$$\begin{aligned}\left\|f(p+v)-f(p)-\sum_{j=1}^{n}(D_pf)(v^ie_i)\right\| &\leq \sum_{j=1}^{n}\left\|f(p+w_{j-1}+v^je_j)-f(p+w_{j-1})-(D_{p+w_{j-1}}f)(v^je_j)\right\| \\ &\quad+\sum_{j=1}^{n}\left\|(D_{p+w_{j-1}}f)(v^je_j)-(D_pf)(v^je_j)\right\| \\ &\leq \sum_{j=1}^{n}\varepsilon\left\|v^je_j\right\|+\sum_{j=1}^{n}\left|v^j\right|\left\|(D_{p+w_{j-1}}f)(e_j)-(D_pf)(e_j)\right\| \\ &\leq 2\varepsilon\left\|v\right\|_\infty \\ &\leq 2C\varepsilon\left\|v\right\|,\end{aligned}$$

the final estimate holding as $\dim V<\infty$ (so that all norms on V are equivalent). This shows that for all nonzero $v\in V$ with $\|v\|$ sufficiently small,

$$\frac{\left\|f(p+v)-f(p)-\sum_{j=1}^{n}(D_pf)(v^ie_i)\right\|}{\|v\|}<\varepsilon.$$

That is, f is differentiable at p with $Df\big|_p(v)=\sum_{i=1}^{n}D_pf(v^ie_i)$.

It remains to show $q\mapsto Df\big|_q$ is continuous. We will show that $q\mapsto Df\big|_q$ is continuous at fixed but arbitrary $p\in U$. Note that for $q\in U$ we have $Df\big|_q\in\mathrm{Hom}(V,W)$ so we need to show that $\left\|Df\big|_q-Df\big|_p\right\|_{\mathrm{op}}\to 0$ as $\|q-p\|\to 0$. Take $q\in B_\delta(p)$ and consider the difference

$$\begin{aligned}\left\|Df\big|_p(v)-Df\big|_q(v)\right\| &= \left\|\sum_{i=1}^{n}v^i\left[(D_pf)(e_i)-(D_qf)(e_i)\right]\right\| \\ &\leq \sum_{i=1}^{n}\left|v^i\right|\left\|(D_pf)(e_i)-(D_qf)(e_i)\right\| \\ &\leq \varepsilon\sum_{i=1}^{n}\left|v^i\right| \\ &\leq \varepsilon C_1\left\|v\right\|.\end{aligned}$$

If $v\neq 0$ divide through by $\|v\|$ to get

$$\frac{\left\|Df\big|_p(v)-Df\big|_q(v)\right\|}{\|v\|}\leq\varepsilon C_1$$

By homogeneity of $\|\cdot\|$ and homogeneity of directional derivatives we get

$$\left\|\left(Df\big|_p-Df\big|_q\right)\left(\frac{v}{\|v\|}\right)\right\|\leq\varepsilon C_1$$

so that
$$\left\|Df\big|_q - Df\big|_p\right\|_{\text{op}} \to 0 \qquad \text{as } q \to p.$$

■

Definition 1.3.6. A subset U of a vector space V is called *convex* if for all $p,q \in U$ and all $t \in [0,1]$, the point $tp+(1-t)q \in U$ (i.e. for all points p and q in U, the segment joining p and q lies entirely in U.

Theorem 1.3.7 (Mean Value Theorem for vector-valued functions)**.** *Suppose U is an open convex subset of a vector space V and $f \in C^1(U)$. If there exists a constant $M > 0$ such that $\left\|Df\big|_q\right\| \leq M$ for all $q \in U$ then*
$$d(f(p),f(q)) \leq Md(p,q) \qquad \text{for all } p,q \in U.$$

Proof. Fix $f \in C^1(U)$. For $p,q \in U$ set $v = q-p$ (so $q = p+v$) and let $S = \{tp+(1-t)q\}_{0\leq t\leq 1}$. Then
$$f(q)-f(p) = \int_0^1 \frac{d}{dt} f(p+tv)\,dt = \int_0^1 (D_{p+tv}f)(v)\,dt.$$
Taking norms on both sides we have
$$\begin{aligned} \|f(q)-f(p)\| &\leq \int_0^1 \left\|(D_{p+tv}f)(v)\right\|\,dt \\ &\leq \int_0^1 \left\|Df\big|_{p+tv}\right\| \|v\|\,dt \\ &\leq \underbrace{\sup_{p'\in S}\left\|Df\big|_{p'}\right\|}_{M} \|v\| \\ &\leq M\|p-q\|. \end{aligned}$$

■

As the next example shows, we can not expect equality in the Mean-Value Theorem for vector-valued functions.

Example 1.3.8. Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $\theta \mapsto (\cos\theta, \sin\theta)$. Choose $p = 0$ and $q = 2\pi$. Then
$$\|f(q)-f(p)\| = \|(1,0)-(1,0)\| = 0.$$
On the other hand, for all $p' \in \mathbb{R}$ (in particular for all p' between $p = 0$ and $q = 2\pi$),
$$\left\|Df\big|_{p'}(q-p)\right\| = \left\|(-\sin(p'),\cos(p'))\cdot 2\pi\right\| = 2\pi.$$

Thus, for all p' between 0 and 2π,

$$\|f(q)-f(p)\| = 0 < 2\pi = \left\| Df\big|_{p'}(q-p) \right\|.$$

△

1.4 Second Derivatives

Throughout this section V and W will be finite-dimensional vector spaces. Let U be an open subset of V and suppose $f : U \to W$ is differentiable. We get a map $U \to \mathrm{Hom}(V,W)$ via $q \mapsto Df\big|_q$. In view of Definition 1.1.1, this map is differentiable at $p \in U$ if there is a linear map $T : V \to \mathrm{Hom}(V,W)$ such that

$$\frac{\left\| Df\big|_{p+h} - Df\big|_p - T(h) \right\|_{\mathrm{op}}}{\|h\|} \to 0 \qquad \text{as } h \to 0.$$

In this case, by Claim 1.1.3, the linear map $T : V \to \mathrm{Hom}(V,W)$ is unique and we will denote this map by $T = D(Df)\big|_p = D_p(Df)$. That is, $D(Df)\big|_p \in \mathrm{Hom}(V,\mathrm{Hom}(V,W))$. Evaluating this map at $v \in V$ gives $(D_p(Df))(v) \in \mathrm{Hom}(V,W)$. Evaluating the map $(D_p(Df))(v)$ at $u \in V$ gives $((D_p(Df))(v))(u) \in W$. This notion of a second derivative is complicated, so some insight on the space $\mathrm{Hom}(V,\mathrm{Hom}(V,W))$ is in order. Define $\mathrm{Bil}(V \times V, W)$ to be the vector space of bilinear maps $V \times V \to W$ with norm

$$\|F\| = \sup_{\|u\|,\|v\|=1} \|F(u,v)\|.$$

For our purposes an important fact is the following:

The vector spaces $\mathrm{Hom}(V,\mathrm{Hom}(V,W))$ *and* $\mathrm{Bil}(V \times V, W)$ *are isometrically isomorphic.*

A proof of this fact can be found in Lemma 2.0.12 of the appendix. The major consequence of this lemma is that for each $p \in U$, we can (naturally and linearly) identify the map $D_p(Df) \in \mathrm{Hom}(V,\mathrm{Hom}(V,W))$ with a map $D_p^2 f \in \mathrm{Bil}(V \times V, W)$ by

$$(D_p^2 f)(u,v) = (D_p(Df)(v))(u).$$

The bilinear mapping $D_p^2 f : V \times V \to W$ is much easier to understand than the linear, $\mathrm{Hom}(V,W)$-valued mapping $D_p(Df)$.

Theorem 1.4.1. *Let $\{e_i\}$ be a basis for V. Assume $(D_p^2 f)(e_i,e_j)$ exists for all p and all i,j and that $p \mapsto (D_p^2 f)(e_i,e_j)$ is continuous. Then* *(a)* $(D_p^2 f)(e_i, e_j) = (D_p^2 f)(e_j, e_i)$ *(equality of mixed partials)*

(b) $(D_p^2 f)(u,v)$ *exists for all* $u, v \in V$.

(c) Bilinearity: For each fixed v, $u \mapsto (D_p^2 f)(u,v)$ *is linear in u. For each fixed u,* $v \mapsto (D_p^2 f)(u,v)$ *is linear in v.*

(d) $D_p^2 f$ *is symmetric:* $(D_p^2 f)(u,v) = (D_p^2 f)(v,u)$ *for all* $u, v \in V$

(e) $(D_p(Df)(v))(u) = (D_p^2 f)(u,v)$. *[Note: If* $f : U \subset V \to W$, *then* $Df : U \to \mathrm{Hom}(V,W)$ *is the map* $p \mapsto D_p f$. *Thus,* $D(Df) : U \to \mathrm{Hom}(V, \mathrm{Hom}(V,W))$ *is the map* $q \mapsto D_q(Df) : V \to \mathrm{Hom}(V,W)$. *Evaluating this map at* $v \in V$ *gives* $(D_q(Df))(v) \in \mathrm{Hom}(V,W)$. *Finally, evaluating the map* $(D_q(Df))(v)$ *at* $u \in V$ *gives* $((D_q(Df))(v))(u) \in W$.*]*

Proof. omitted. ■

Example 1.4.2. Let $V = \{n \times n \text{ matrices}\} = M_{n\times n}(\mathbb{R}) = W$. So $\dim(V) = n^2 = \dim(W)$. the standard basis is $\{e_{ij}\}$, where e_{ij} is the matrix whose (ij)th entry is 1 and all other entries are zero. Define $f : V \to V$ to be the squaring map

$$f(A) = A^2.$$

Let $A, B \in M_{n\times n}(\mathbb{R})$ Then

$$(D_A f)(B) = \frac{d}{dt}(A+tB)^2\Big|_{t=0} = \frac{d}{dt}\left(A^2 + t(AB+BA) + t^2B^2\right)\Big|_{t=0} = AB + BA.$$

Remark 1.4.3. If we fix a basis, we can get a Jacobian matrix $n^2 \times n^2$, however this point will not be pursued here.

For fixed B, the map $A \mapsto AB + BA$ is continuous. Therefore, f is differentiable and $Df\big|_A : M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ is given by $Df\big|_A(B) = AB + BA$.
Compute$(D_A^2 f)(B,C)$:

$$(D_A(Df))(B)(C) = (D_A^2 f)(B,C) = D_A(\underbrace{A' \mapsto (D_{A'} f)(B)}_{A'B+BA'})(C) = CB + BC.$$

Notice that this computation illustrates the fact that “the derivative of a liner map is itself” (see Example 1.1.5). So, $(D_A^2 f)(B,C) = BC + CB$ is constant as a function of A. Conclude that $D^3 f = 0$.

Example 1.4.4. Let $V = M_{n\times n}(\mathbb{R}) = W$ and let $U = GL_n(\mathbb{R}) = \{A \in M_{n\times n}(\mathbb{R}) : A \text{ is invertible}\}$. Note that U is an open subset of V as U is the preimage of the open set $\mathbb{R}\setminus\{0\}$ under the continuous function $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ (you can check on your own that the determinant is continuous).

In this example we will compute the first two derivatives of the inversion map $i : GL_n(\mathbb{R}) \to GL_n(\mathbb{R})$ given by $i(A) = A^{-1}$. As usual, in order to guess the formula for $Di\big|_A$, we should consider $D_A i(B)$, the directional derivative of i at A in the direction of B. We have

$$D_A i(B) = \lim_{t\to 0} \frac{i(A+tB) - i(A)}{t},$$

whenever the limit exists. Since the numerator in this quotient is equal to

$$(A+tB)^{-1} - A^{-1} = (A(I+tA^{-1}B))^{-1} - A^{-1} = ((I+tA^{-1}B)^{-1} - I)A^{-1},$$

we see that

$$D_A i(B) = D_I i(A^{-1}B)\cdot A^{-1}. \tag{1.5}$$

In order to make our lives easier, let us compute $D_I i(B)$ for $B \in GL_n(\mathbb{R})$. Once we have computed $D_I i(B)$, we can use this result with B replaced by $A^{-1}B$ in equation (1.5) to recover $D_A i(B)$. To compute $D_I i(B)$ we need to compute $\frac{d}{dt}(I+tB)^{-1}\big|_{t=0}$. We achieve this with a "geometric series trick".

Claim 1.4.5. *If $A \in M_{n\times n}(\mathbb{R})$ with $\|A\| < 1$ then $S_m = \sum_{j=0}^{m} A^j$ converges (with respect to $\|\cdot\|_{\text{op}}$) to $(I-A)^{-1}$.*

Proof of Claim. first observe that for any integer $j \geq 1$ and any $x \in \mathbb{R}^n$ with $\|x\| = 1$,

$$\left\|A^j x\right\| = \left\|AA^{j-1}x\right\| \leq \|A\| \left\|A^{j-1}x\right\|.$$

Repeating this estimate $j-1$ more times gives

$$\left\|A^j x\right\| \leq \|A\|^j \|x\| = \|A\|^j,$$

where $\|x\| = 1$ was used in the final equality. Taking the supremum over all $\|x\| = 1$ gives $\left\|A^j\right\| \leq \|A\|^j$.
Next, we show that S_m converges to some (bounded) linear operator as $m \to \infty$. Since $\text{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ is complete, it suffices to show that S_m is a Cauchy sequence. Accordingly, let $m, k \in \mathbb{N}$. We have

$$\begin{aligned}\|S_{m+k} - S_k\| &= \left\|A^{k+1} + \cdots + A^{k+m}\right\| \\ &\leq \left\|A^{k+1}\right\| \|S_{m-1}\| \\ &\leq \|A\|^{k+1} \|S_{m-1}\|.\end{aligned}$$

Now, for any $m \in \mathbb{N}$, by comparing to the geometric series $\sum_{j=0}^{\infty} \|A\|^j$, we have $\|S_{m-1}\| \leq (1-\|A\|)^{-1}$. Therefore, letting $k,m \to \infty$ in the above string of inequalities gives $\|S_{m+k} - S_k\| \to 0$ as $k,m \to \infty$.
Finally, we show that $S_\infty := \sum_{j=0}^{\infty} A^j$ satisfies $S_\infty = (I-A)^{-1}$. For any $m \in \mathbb{N}$, we have $(I-A)S_m = I - A^{m+1}$, so

$$(I-A)S_\infty - I = -A^{m+1} - (I-A)(S_m - S_\infty).$$

After taking norms we get

$$\|(I-A)S_\infty - I\| \leq \|A\|^{m+1} + \|I-A\| \, \|S_m - S_\infty\| \, .$$

Letting $m \to \infty$ in this estimate shows that S_∞ is a right-inverse of $I-A$. By a similar argument we have that S_∞ is a left-inverse of $I-A$. The claim is estiablished. ■

Now, using the claim, we have for $|t|$ sufficiently small ($|t| \, \|B\| < \frac{1}{2}$ is sufficient),

$$(I+tB)^{-1} = (I-(-tB))^{-1} = \sum_{k=0}^{\infty} (-1)^k t^k B^k$$

so

$$D_I i(B) = \frac{d}{dt} (I+tB)^{-1} \big|_{t=0} = \frac{d}{dt} \sum_{k=0}^{\infty} (-1)^k t^k B^k \big|_{t=0} = \sum_{k=1}^{\infty} (-1)^k k t^{k-1} B^k \big|_{t=0} = -B.$$

(Note that the term-by-term differentiated series $\sum_{k=1}^{\infty} (-1)^k k t^{k-1} B^k$ converges uniformly for $|t|$ sufficiently small). Replacing B by $A^{-1}B$ in this equality and in view of equation (1.5) we have for $A, B \in GL_n(\mathbb{R})$,

$$D_A i(B) = D_I i(A^{-1}B) \cdot A^{-1} = -A^{-1}BA^{-1}.$$

Since, for fixed $B \in GL_n(\mathbb{R})$ the map $A \mapsto -A^{-1}BA^{-1}$ is continuous, Proposition 1.3.3 implies that i is differentiable and $Di\big|_A : GL_n(\mathbb{R}) \to GL_n(\mathbb{R})$ is given by

$$Di\big|_A(B) = D_A i(B) = -A^{-1}BA^{-1}.$$

This is the matrix-version of the first-semester calculus result $\frac{d}{dx}(x^{-1}) = -x^{-2}$ for $x \in \mathbb{R}$.

Now, for $A, B, C \in GL_n(\mathbb{R})$ let us compute

$$D_A^2 i(B,C) = D_A \left((M \mapsto Di\big|_M)(B) \right) (C),$$

the directional derivative of the map $(M \mapsto Di\big|_M)(B) : GL_n(\mathbb{R}) \to GL_n(\mathbb{R})$ at A in direction C. We have

$$\begin{aligned} D_A\left((M \mapsto Di\big|_M)(B)\right)(C) &= \frac{d}{dt}\left(-(A+tC)^{-1}B(A+tC)^{-1}\right)\Big|_{t=0} \\ &= -\left[-A^{-1}CA^{-1}BA^{-1} - A^{-1}BA^{-1}CA^{-1}\right] \\ &= A^{-1}CA^{-1}BA^{-1} + A^{-1}BA^{-1}CA^{-1}. \end{aligned}$$

This is the matrix version of the first-semester calculus result $\frac{d^2}{dx^2}\left(x^{-1}\right) = 2x^{-3}$.

Similarly, one can compute for $A,B,C,D \in GL_n(\mathbb{R})$,

$$D_A i(B,C,D) = \sum (\text{six terms}),$$

which is the matrix version of the first-semester calculus result $\frac{d^3}{dx^3}\left(x^{-1}\right) = 3!\, x^{-4}$.

1.5 Chain Rule

Throughout this section V, W and X will be finite-dimensional vector spaces and $A \subset V$ and $B \subset W$ will be open sets. Let $f : V \to W$ and $g : W \to X$ and suppose $p \in A$ with $f(p) \in B$. The best statement of the chain rule is

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

Theorem 1.5.1. *With data as above, if f is differentiable at p and g is differentiable at $f(p)$ then $g \circ f$ is differentiable at p and*

$$D(g \circ f)\big|_p = Dg\big|_{f(p)} \circ Df\big|_p.$$

Proof. The proof is left as an exercise. One should use the definition of the derivative. ■

Consider the special case where $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ and $X = \mathbb{R}^k$. In this case, as in Example 1.1.9 we have

$$Df\big|_p(v) = \underbrace{J_f(p)}_{m\times n}\underbrace{v}_{\in R^n} \in \mathbb{R}^m \qquad \text{and} \qquad Dg\big|_q(w) = \underbrace{J_g(q)}_{k\times m}\underbrace{w}_{\in R^m} \in \mathbb{R}^k.$$

Theorem 1.5.1 implies

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

Therefore, (the second-best statement of the chain rule)

$$J_{g\circ f}(p) = J_g(f(p))J_f(p).$$

Corollary 1.5.2. *In the setting of Theorem 1.5.1, if f and g are C^1 then so is $g \circ f$.*

Proof. Take a look at the the diagram in Figure 1.1. Since each of these maps is continu-

$$A \longrightarrow A \times B \longrightarrow \mathrm{Hom}(V,W) \times \mathrm{Hom}(W,X) \longrightarrow \mathrm{Hom}(V,X)$$

$$p \longmapsto (p, f(p))$$

$$(p,q) \longmapsto (Df\big|_p, Dg\big|_q)$$

$$(T,S) \longmapsto (S \circ T)$$

$$p \longmapsto Dg\big|_{f(p)} \circ Df\big|_p$$

Figure 1.1: Writing the map $p \mapsto Dg\big|_{f(p)} \circ Df\big|_p$ as the composition of continuous maps

ous so the map $A \to \mathrm{Hom}(V,X)$ given by

$$p \mapsto Dg\big|_{f(p)} \circ Df\big|_p$$

is continuous. ■

Corollary 1.5.3. *Let $1 \le k \le \infty$. If f and g are composable C^k maps then $g \circ f$ is also C^k.*

Proof. Use induction and the "composition trick" from Corollary 1.5.2. ■

1.6 Inverse and Implicit Function Theorems

Throughout this section, X and Y will denote finite-dimensional vector spaces and U will denote an open subset of X.

Definition 1.6.1. A map g is called a *diffeomorphism* if g is a differentiable map with differentiable inverse. If both g and g^{-1} are C^k we call g a C^k-diffeomorphism. Say g is a *local diffeomorphism at p.* if there exists a (small) neighborhood $U \subset \text{domain}(g)$ with $p \in U$ such that $g\big|_U$ is a diffeomorphism. Say g is a *local diffeomorphism* if g is a local diffeomorphism at each $p \in \text{domain}(g)$.

Theorem 1.6.2 (Inverse Function Theorem)**.** *Let X and Y be finite-dimensional vector spaces, let $U \subset X$ be open and let $f : U \to Y$ be a C^1 map. If $x_0 \in U$ is such that $D_{x_0} f$ is invertible then f is a local C^1-diffeomorphism at x_0. That is, if $D_{x_0} f$ is invertible, then there exists open neighborhoods U_1 of x_0 and V_1 of $f(x_0)$ such that f maps U_1 to V_1 bijectively and such that*

$$\left(f\big|_{U_1}\right)^{-1} : V_1 \to U_1$$

is C^1.

Proof. omitted. ■

As the next corollary shows, a simple application of the chain rule yields an expression for the derivative of f^{-1} in terms of the derivative of f.

Corollary 1.6.3 (Corollary to Inverse Function Theorem)**.** *In the setting of Theorem 1.6.2,*

$$D_y f^{-1} = \left(D_{f^{-1}(y)}\right)^{-1}$$

for $y \in V_1$. Here, f^{-1} is the inverse of $f\big|_{U_1} : U_1 \to V_1$.

Proof. Let $h = \left(f\big|_{U_1}\right)^{-1}$. Then $f \circ h = \text{id}_{V_1} = \text{id}_Y\big|_{V_1}$. Note that $f \circ h$ is linear so $f \circ h$ is "its own derivative". By the chain rule (applied to the function $f \circ h$),

$$\text{id}_Y = D_y(f \circ h) = D_{h(y)} f \circ D_y h.$$

Therefore, $D_y h = \left(D_{h(y)} f\right)^{-1}$. ■

Corollary 1.6.4. *In the setting of the Inverse Function Theorem, if f is C^k, where $1 \le k \le \infty$, then so is the locally-defined $f^{-1} = h$.*

Sketch of Proof. The idea of the (beginning of the) proof is that if $f \in C^2$ then one can to use the Inverse Function Theorem together with Corollary 1.6.3 to write $y \mapsto D_y f^{-1}$ as a composition of C^1 maps. Observe that $Dh : V_1 \to \text{Hom}(Y, X)$ is given by

$$y \mapsto D_y h = \left(D_{f^{-1}(y)} f\right)^{-1} = (i \circ Df \circ f^{-1})(y),$$

$$\begin{array}{ccccccc} V_1 & \longrightarrow & U_1 & \xrightarrow{Df} & \mathrm{Hom}(X,Y) & \longrightarrow & \mathrm{Hom}(Y,X) \\ y & \longmapsto & h(y) & & T & \xmapsto[\text{a } C^\infty \text{ map}]{} & T^{-1} \end{array}$$

Figure 1.2: If $f \in C^k$, then Dh is a composition of C^{k-1} maps.

where

$$i : (\text{open subset of} \, \mathrm{Hom}(X,Y)) \to \mathrm{Hom}(Y,X)$$

is the inversion map. Assume f is C^2. Then Df is C^1. Moreover, f^{-1} is C^1 (by the Inverse Function Theorem) and i is C^1 (see Example 1.4.4 for the idea behind a proof of this fact). Therefore, Dh is C^1 and h is C^2. Now induct on k. ■

Motivation for the Implicit Function Theorem. Suppose we have a system of m scalar equations in $m+n$ unknowns ($n > 0$). Write the system as follows:

$$\begin{aligned} &f^1(x,y) = 0 \\ &f^2(x,y) = 0 \\ &\vdots \\ &f^m(x,y) = 0, \end{aligned}$$

were $x = (x^1, \cdots, x^n)$ and $y = (y^1, \cdots, y^m)$. Want to solve for $\{y^i\}$ in terms of $\{x^j\}$. Morally we should be able to use each equation to eliminate one variable iteratively:
Use the first equation to write

$$y^m = \text{ function of } (x, y^1, y^2, \cdots, y^{m-1}).$$

Substitute this into remaining equations to get a new system of $m-1$ equations in $m+n-1$ unknowns:

$$\begin{aligned} &\tilde{f}^{(2)}(x, y^1, \cdots, y^{m-1}) = 0 \\ &\vdots \\ &\tilde{f}^{(m)}(x, y^1, \cdots, y^{m-1}) = 0. \end{aligned}$$

Use the new first equation to solve for y^{m-1} in terms of the remaining variables to get

$$y^{m-1} = \text{ function of}(x, y^1, \cdots, y^{m-2}).$$

This expression for y^{m-1} can also be used to express y^m as a function of $(x, y^1, \cdots, y^{m-2})$. Repeat as necessary to end up with $y = (y^1, \cdots, y^m)$ in terms of x.

For notational convenience we define the vector-valued function $F : X \times Y \to W$ by

$$F(x,y) = \begin{pmatrix} f^1(x,y) \\ \vdots \\ f^m(x,y) \end{pmatrix}$$

so that the system

$$\begin{cases} f^1(x,y) = 0 \\ \quad\vdots \\ f^m(x,y) = 0 \end{cases}$$

becomes $F(x,y) = 0$.

$$\begin{array}{ccc} X \times Y & \xrightarrow{\quad F \quad} & W \\ \cup \quad \cup & & \\ A \quad B & & \end{array}$$

Write elements of $X \times Y$ as either (x,y) (i.e. points of $A \times B$) or as $\begin{pmatrix} x \\ y \end{pmatrix}$ (i.e. as a vector to which a derivative is applied)

Temporarily, for $i = 1,2$, let $D_p^{[i]}F$ for the linear map

$$\begin{cases} X \to W & \text{if } i = 1 \\ Y \to W & \text{if } i = 2 \end{cases}$$

obtained by differentiating with respect to the i^{th} factor of $X \times Y$ holding variable in other factor fixed e.g.

$$(D_p^{[2]}F)(v) = (D_pF)\begin{pmatrix} 0 \\ v \end{pmatrix}$$

Theorem 1.6.5 (Implicit Function Theorem)**.** *With data as in above diagram, assume $F : A \times B \to W$ is C^1. Let $(x_0,y_0) \in A \times B$ and assume $F(x_0,y_0) = 0$. Suppose $D_{(x_0,y_0)}^{[2]}F : Y \to W$ is invertible. Then there exists open neighborhoods A_1 of x_0 and B_1 of y_0 and a C^1 function $g : A_1 \to B_1$ such that for all $(x,y) \in A_1 \times B_1$, $F(x,y) = 0$ if and only if $y = g(x)$. See Figure 1.3.*

Under the hypotheses of the setting of the Implicit Function Theorem, the level set of F locally defines y as a C^1 function of x. The next corollary shows us how to compute the derivative of this function in terms of F.

Corollary 1.6.6 (Corollary to Implicit Function Theorem)**.** *In the setting of Theorem 1.6.5,*

$$D_xg = -\left(D_{(x,g(x))}^{[2]}F\right)^{-1} \circ D_{(x,g(x))}^{[1]}F.$$

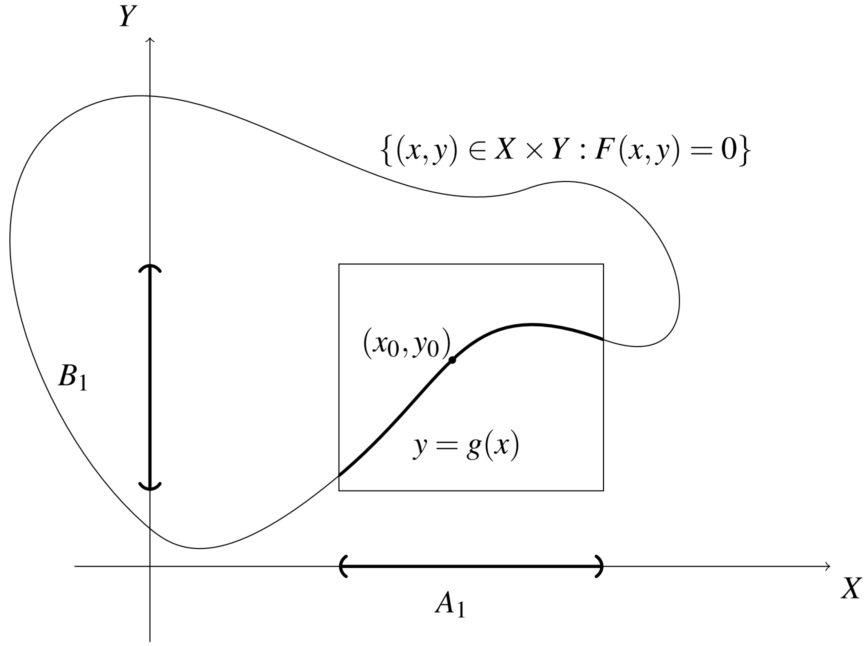

Figure 1.3: Figure for Implicit Function Theorem. If (x_0, y_0) is on the level-set of F and if $D^{[2]}_{(x_0,y_0)}F$ is invertible then the level set of F locally defines y as a C^1 function of x.

The proof of Corollary 1.6.6 is omitted but follows routinely from the following lemma.

Lemma 1.6.7. *The derivative of the map* $h : x \mapsto F(x, g(x))$ *is given by*

$$D_x h = D^{[1]}_{(x,g(x))}F + \left(D^{[2]}_{(x,g(x))}F\right) \circ D_x g.$$

Proof of Lemma. h is a composition as in Figure 1.4. Now you can finish the proof.

$$A_1 \longrightarrow A_1 \times B_1 \longrightarrow W$$

$$x \longmapsto (x, g(x)) \longmapsto F(x, g(x))$$

Figure 1.4: h is a composition

■

Note: If x and y are one-dimensional variables and $F(x,y)=0$ then differentiability with respect to x gives

$$\frac{\partial F}{\partial x}(x,y)+\frac{\partial F}{\partial y}(x,y)\frac{dy}{dx}=0$$

Solving for $\frac{dy}{dx}$ yields

$$\frac{dy}{dx}=-\frac{\partial F}{\partial x}\left(\frac{\partial F}{\partial y}\right)^{-1},$$

the usual formula for "implicit differentiation" from first-semester calculus.

Corollary 1.6.8 (Corollary to Implicit Function Theorem)**.** *The g given by the Implicit Function Theorem is as continuously differentiable as the F.*

Proof. Do it on your own. ■

1.6.1 Equivalence of the Inverse and Implicit Function Theorems

This section has not been proof read

Claim 1.6.9. *The Implicit Function Theorem implies the Inverse Function Theorem.*

Proof. Assume the Implicit Function Theorem holds and assume the hypotheses of the Inverse Function Theorem. Define $F: Y\times U\to Y$ by

$$F(y,x)=y-f(x) \qquad \text{for } x\in U, y\in Y.$$

It is easy to show that F is C^1 since f is. Let $y_0=f(x_0)$ so that $F(y_0,x_0)=0$.

$$D^{[2]}_{(x_0,y_0)}F=-D_{x_0}f$$

is invertible. By the Implicit Function Theorem, there exists neighborhoods A_1 of y_0 and B_1 of x_0 and a C^1 map $g: A_1\to B_1$ such that for all $(y,x)\in A_1\times B_1$, $F(y,x)=0$ if and only if $x=g(y)$ (i.e. $y=f(x)$ if and only if $x=g(y)$). So

$$g=\left(f|_{B_1}\right)^{-1}.$$

■

Claim 1.6.10. *The Inverse Function Theorem Implies the Implicit Function Theorem.*

Proof. Assume the Inverse Function Theorem holds and assume the hypotheses of the Implicit Function Theorem. Have

$$F : \underbrace{A}_{\dim n} \times \underbrace{B}_{\dim m} \to \underbrace{W}_{\dim m}$$

F can be invertible since the dimensions of the domain of F and the codomain of F do not coincide. Define $f : A \times B \to A \times W$ by

$$(x,y) \mapsto (x, F(x,y))$$

(so the dimension of f's domain and the dimension of f's codomain coincide). Then f is C^1 and

$$\begin{aligned}
(D_{(x,y)}f)\begin{pmatrix} a \\ 0 \end{pmatrix} &= \frac{d}{dt} f \begin{pmatrix} x+ta \\ y \end{pmatrix}\Bigg|_{t=0} \\
&= \frac{d}{dt} \begin{pmatrix} x+ta \\ F(x+ta,y) \end{pmatrix}\Bigg|_{t=0} \\
&= \begin{pmatrix} a \\ (D^{[1]}_{(x,y)}F)(a) \end{pmatrix}.
\end{aligned}$$

Similarly,

$$(D_{(x,y)}f)\begin{pmatrix} 0 \\ b \end{pmatrix} = \begin{pmatrix} a \\ (D^{[2]}_{(x,y)}F)(b) \end{pmatrix}.$$

Take $(x,y) = (x_0, y_0)$. For notational convenience, define linear maps

$$S = D^{[1]}_{(x_0,y_0)}F \qquad \text{and} \qquad T = D^{[2]}_{(x_0,y_0)}F.$$

Then

$$\begin{aligned}
(D_{(x_0,y_0)})\begin{pmatrix} a \\ b \end{pmatrix} &= (D_{(x_0,y_0)})\begin{pmatrix} a \\ 0 \end{pmatrix} + (D_{(x_0,y_0)})\begin{pmatrix} 0 \\ b \end{pmatrix} \\
&= \begin{pmatrix} a \\ S(a) \end{pmatrix} + \begin{pmatrix} 0 \\ T(b) \end{pmatrix} \\
&= \begin{pmatrix} a \\ S(a)+T(b) \end{pmatrix}.
\end{aligned}$$

Therefore, $D_{(x_0,y_0)}f$ is the map

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} a \\ S(a)+T(b) \end{pmatrix}.$$

Next, we show this map is invertible. Accordingly, suppose

$$\begin{pmatrix} a \\ S(a)+T(b) \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}.$$

Then $a = c$ and $b = T^{-1}(d - S(c))$ so the inverse exists and is given by

$$\begin{pmatrix} c \\ d \end{pmatrix} \mapsto \begin{pmatrix} a \\ T^{-1}(d - S(c)) \end{pmatrix}.$$

Since $D_{(x_0,y_0)}f$ is invertible, the Inverse Function Theorem can be applied. We obtain an open neighborhood U_1 of $(x_0,y_0) \in A_1 \times B_1 \subset X \times Y$ and an open neighborhood V_1 of $f(x_0,y_0) = (x_0, F(x_0,y_0)) = (x_0,0) \in A \times W \subset X \times W$ such that

$$f\big|_{U_1} : U_1 \to V_1$$

is a C^1 diffeomorphism.

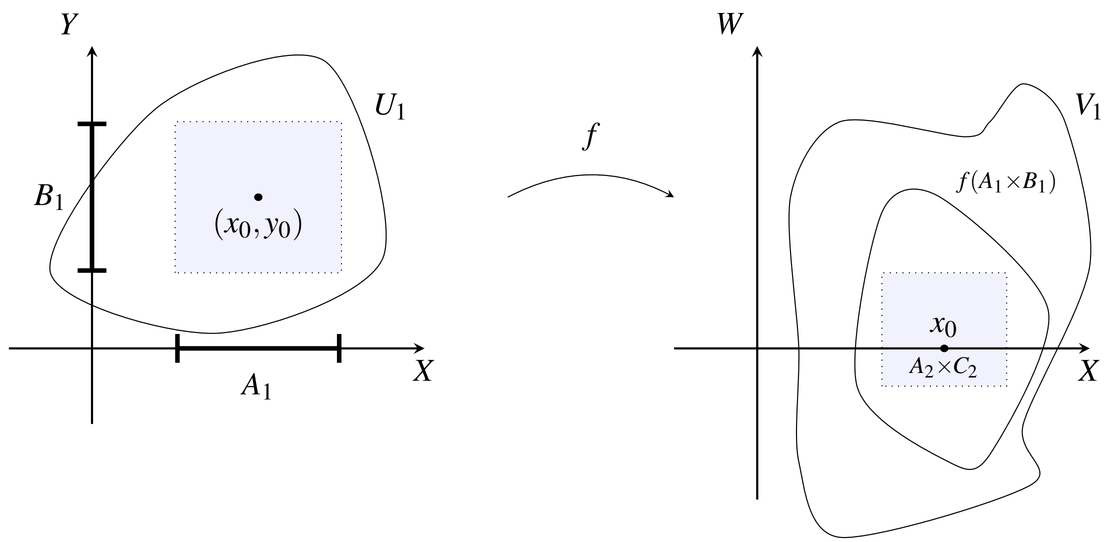

Figure 1.5: Put the figure

Since U_1 is open, choose open neighborhoods A_1 of x_0 and B_1 of y_0 such that $A_1 \times B_1 \subset U_1$. Since f is an open, so is $f(A_1 \times B_1)$. Choose open neighborhoods A_2 of x_0 and C_2 of $0 \in W$ such that $A_2 \times C_2 \subset f(A_1 \times B_2)$. Set $\tilde{g} = \left(f\big|_{A_1 \times B_1}\right)^{-1}$, $\tilde{g} : f(A_1 \times B_1) \to A_1 \times B_1$. Write

$$\tilde{g}(x,w) = (g_1(x,w), g_2(x,w)) \qquad \text{for } x,w \in f(A_1 \times B_1)$$

f is a C^1 diffeomorphism, so $\tilde{g}$ is C^1. Moreover, so are the component functions g_1 and g_2. We have

$$f \circ \tilde{g} = \text{id} \qquad \text{on } A_2 \times C_2$$

(this is true on a larger set, but we only care about $A_2 \times C_2$). So, for all $(x,w) \in A_2 \times C_2$,

$$\begin{aligned}(x,w) &= f(g_1(x,w),g_2(x,w)) \\ &= (g_1(x,w),F(g_1(x,w),g_2(x,w))).\end{aligned}$$

Therefore, $x = g_1(x,w)$ and

$$w = F(g_1(x,w),g_2(x,w)) = F(x,g_2(x,w)).$$

Taking $w = 0$ shows that

$$F(x,g_2(x,0)) = 0 \qquad \text{for all } x \in A_2.$$

Define $\hat{g} : A_2 \to B_1$ by $\hat{g}(x) = g_2(x,0)$. Then $F(x,\hat{g}(x)) = 0$ for all $x \in A_2$. In other words, if $(x,y) \in A_2 \times B_1$ and $y = \hat{g}(x)$, then $F(x,y) = 0$.

Now, consider $\tilde{g} \circ f = \mathrm{id}$ on $A_1 \times B_1$. For all $(x,y) \in A_1 \times B_1$,

$$\begin{aligned}(x,y) &= \tilde{g}(f(x,y)) \\ &= \tilde{g}(x,F(x,y)) \\ &= (g_1(x,F(x,y)),g_2(x,F(x,y))) \\ &= (x,g_2(x,F(x,y))).\end{aligned}$$

Implies

$$g_2(x,F(x,y)) = y \qquad \text{for all } (x,y) \in A_1 \times B_1.$$

Taking (x,y) to satisfy $F(x,y) = 0$, we have $g_2(x,0) = y$. Set $A_3 = A_1 \cap A_2$ and define $g(x) = \hat{g}(x)$. Then g is C^1 and for all $x \in A_3$, $y \in B_1$ we have

$$y = g(x) \qquad \text{whenever} \qquad F(x,y) = 0.$$

Therefore, for all $x,y \in A_2 \times B_1$, $F(x,y) = 0$ if and only if $y = g(x)$. ■

Chapter 2

Appendix

Proposition 2.0.11. *If V is a vector space of finite dimension n then all norms on V are equivalent.*

The proof will be given for $V = \mathbb{R}^n$. By a standard argument, one can obtain the conclusion of the proposition for general n-dimensional vector spaces.

Proof. Let $\{e_j\}_{j=1}^n$ be the standard basis for $\mathbb{R}^n$. It suffices to show that every norm on $\mathbb{R}^n$ is equivalent to the norm $\|\cdot\|_\infty$ given by $\left\|\sum_{j=1}^n x^j e_j\right\|_\infty = \max_j \left|x^j\right|$. Accordingly, let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$ using the triangle inequality and the homogeneity of $\|\cdot\|$ we have

$$\|x\| = \left\|\sum_{j=1}^n x^j e_j\right\| \le \sum_{j=1}^n \left\|x^j e_j\right\| = \sum_{j=1}^n \left|x^j\right| \left\|e_j\right\| \le n \max_j \left\|e_j\right\| \|x\|_\infty . \tag{2.1}$$

It remains to show that there is a constant $C_1 > 0$ such that for all $x \in \mathbb{R}^n$, the estimate

$$C_1 \|x\|_\infty \le \|x\| \tag{2.2}$$

holds. Define $f : V \to \mathbb{R}$ by $f(x) = \|x\|$. By the triangle inequality and using equation (2.1) we have

$$|\|x\| - \|y\|| \le \|x - y\| \le C \|x - y\|_\infty ,$$

where $C = n \max_j \left\|e_j\right\|$ as in (2.1). This estimate says that f is continuous from V with the $\|\cdot\|_\infty$-topology to $\mathbb{R}$ with the usual topology. Since $\dim V = n < \infty$, the unit sphere $S^{n-1} = \{x \in V : \|x\|_\infty = 1\}$ is compact. Therefore, f attains its minimum value over S^{n-1} at some point of S^{n-1}. That is, there is $x_0 \in S^{n-1}$ such that $\|x_0\| = \min_{x \in S^{n-1}} \|x\|$. Finally, if $x \neq 0$ we have

$$\|x\| = \|x\|_\infty \left\| \frac{x}{\|x\|_\infty} \right\| \ge \|x_0\| \|x\|_\infty ,$$

while the estimate $\|x\| \ge \|x_0\| \|x\|_\infty$ holds trivially for $x = 0$. Thus, we have established estimate (2.2) with $C_1 = \|x_0\|$. ■

Lemma 2.0.12. *The vector spaces* $\mathrm{Hom}(V, \mathrm{Hom}(V,W))$ *and* $\mathrm{Bil}(V \times V, W)$ *are isometrically isomorphic.*

Proof. Define the map $\phi : \mathrm{Hom}(V, \mathrm{Hom}(V,W)) \to \mathrm{Bil}(V \times V, W)$ by

$$(\phi(T))(u,v) = (T(v))(u) \qquad \text{for } u, v \in V.$$

First we show that $\mathrm{Range}(\phi) \subset \mathrm{Bil}(V \times V, W)$. Fix $T \in \mathrm{Hom}(V, \mathrm{Hom}(V,W))$. For $a, b \in \mathbb{R}$ and $u, v \in V$,

$$\begin{aligned} \phi(T)(au, bv) &= (T(bv))(au) \\ &= a(T(bv))(u) \\ &= ab(T(v))(u) \\ &= ab\phi(T)(u,v). \end{aligned}$$

the second equality holding as $T(bv) \in \mathrm{Hom}(V,W)$ and the third equality holding by the linearity of T. Similarly, for $u_i, v_i \in V$ $(i = 1,2)$ we have

$$\begin{aligned}
\phi(T)(u_1+u_2, v_1+v_2) &= (T(v_1+v_2))(u_1+u_2) \\
&= (T(v_1+v_2))(u_1) + (T(v_1+v_2))(u_2) \\
&= (T(v_1))(u_1) + (T(v_2))(u_1) + (T(v_1))(u_2) + (T(v_2))(u_2) \\
&= \phi(T)(u_1,v_1) + \phi(T)(u_1,v_2) + \phi(T)(u_2,v_1) + \phi(T)(u_2,v_2).
\end{aligned}$$

Next we show that ϕ is linear. Fix $a \in \mathbb{R}$ and $S, T \in \mathrm{Hom}(V, \mathrm{Hom}(V,W))$. For $u, v \in V$ we have

$$\begin{aligned}
(\phi(aS+T))(u,v) &= ((aS+T)(v))(u) \\
&= (aS(v) + T(v))(u) \\
&= a(S(v))(u) + (T(v))(u) \\
&= a\phi(S)(u,v) + \phi(T)(u,v).
\end{aligned}$$

Next we show that ϕ is injective. Suppose $T \in \mathrm{Hom}(V, \mathrm{Hom}(V,W))$ and that $\phi(T)$ is the zero map $V \times V \to W$. Then for all $u, v \in V$,

$$0 = \phi(T)(u,v) = (T(v))(u).$$

Thus, for all $v \in V$, $T(v)$ is the zero map $V \to W$. We conclude that T is the zero map $V \to \mathrm{Hom}(V,W)$.

Next, surjectivity of ϕ follows as $\mathrm{Hom}(V, \mathrm{Hom}(V,W))$ and $\mathrm{Bil}(V \times V, W)$ have the same dimension.

It remains to show that $\|T\| = \|\phi(T)\|$ for all $T \in \mathrm{Hom}(V, \mathrm{Hom}(V,W))$. We have

$$\begin{aligned}
\|\phi(T)\| &= \sup_{\|v\|,\|u\|=1} \|\phi(T)(u,v)\| \\
&= \sup_{\|v\|,\|u\|=1} \|(T(v))(u)\| \\
&= \sup_{\|v\|=1} \left(\sup_{\|u\|=1} \|(T(v))(u)\| \right) \\
&= \sup_{\|v\|=1} \|T(v)\| \\
&= \|T\|.
\end{aligned}$$

■