Elementary Analysis

Joe Erickson

Contents

10 Differentiation in Euclidean Space **1**

10.1 Vector Spaces . . . 1
10.2 Linear Mappings . . . 9
10.3 The Norm of a Linear Mapping . . . 18
10.4 The Total Derivative . . . 23
10.5 Partial Derivatives . . . 27
10.6 The Inverse Function Theorem . . . 36
10.7 The Implicit Function Theorem and Rank Theorem . . . 40
10.8 Derivatives of Higher Order . . . 48
10.9 Differentiation of Integrals . . . 48

10

Differentiation in Euclidean Space

10.1 – Vector Spaces

Most of the linear algebra results given in this section and the next are established in the *Linear Algebra Notes* ([LAN]). Almost all of these results are proven in these pages, but some have proof omitted and the reader is referred to the aforementioned notes. All needed definitions and notations are given here, for the sake of convenience.

We begin with a general definition of the algebraic structure known as a vector space, and later specialize to the setting needed for the analytical developments to come.

Definition 10.1. *A* ***vector space*** *over a field* $\mathbb{F}$ *is a set* V *of objects, along with operations* ***vector addition*** $V \times V \to V$ *(denoted by* $+$*) and* ***scalar multiplication*** $\mathbb{F} \times V \to V$ *(denoted by* $\cdot$ *or juxtaposition) subject to the following axioms:*

VS1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ *for any* $\mathbf{u}, \mathbf{v} \in V$
VS2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ *for any* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$
VS3. *There exists some* $\mathbf{0} \in V$ *such that* $\mathbf{u} + \mathbf{0} = \mathbf{u}$ *for any* $\mathbf{u} \in V$
VS4. *For each* $\mathbf{u} \in V$ *there exists some* $-\mathbf{u} \in V$ *such that* $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
VS5. *For any* $a \in \mathbb{F}$ *and* $\mathbf{u}, \mathbf{v} \in V$, $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
VS6. *For any* $a, b \in \mathbb{F}$ *and* $\mathbf{u} \in V$, $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$
VS7. *For any* $a, b \in \mathbb{F}$ *and* $\mathbf{u} \in V$, $a(b\mathbf{u}) = (ab)\mathbf{u}$
VS8. *For all* $\mathbf{u} \in V$, $1\mathbf{u} = \mathbf{u}$

The elements of V *are called* ***vectors*** *and the elements of the underlying field* $\mathbb{F}$ *are called* ***scalars****.*

We now give some additional properties that, while seemingly obvious, nonetheless require use of many of the axioms listed above to establish.

Proposition 10.2. *Let* V *be a vector space,* $\mathbf{u} \in V$*, and* $a \in \mathbb{F}$*. Then the following hold.*

1. $0\mathbf{u} = \mathbf{0}$.
2. $a\mathbf{0} = \mathbf{0}$.
3. *If* $a\mathbf{u} = \mathbf{0}$*, then* $a = 0$ *or* $\mathbf{u} = \mathbf{0}$.
4. $(-1)\mathbf{u} = -\mathbf{u}$.

Proof.
Proof of Part (1). Since $\mathbf{u} \in V$ and $0 \in \mathbb{F}$, we have $0\mathbf{u} \in V$ by the closure property. Now,

$$\begin{aligned}
0\mathbf{u} &= 0\mathbf{u} + \mathbf{0} && \text{Axiom VS3} \\
&= 0\mathbf{u} + [\mathbf{u} + (-\mathbf{u})] && \text{Axiom VS4} \\
&= (0\mathbf{u} + \mathbf{u}) + (-\mathbf{u}) && \text{Axiom VS2} \\
&= (0\mathbf{u} + 1\mathbf{u}) + (-\mathbf{u}) && \text{Axiom VS8} \\
&= (0 + 1)\mathbf{u} + (-\mathbf{u}) && \text{Axiom VS6} \\
&= 1\mathbf{u} + (-\mathbf{u}) && \\
&= \mathbf{u} + (-\mathbf{u}) && \text{Axiom VS8} \\
&= \mathbf{0}, && \text{Axiom VS4}
\end{aligned}$$

where of course $0 + 1 = 1$ is a known property of real numbers.

The proofs of parts (2), (3), and (4) are left to the exercises. ∎

Definition 10.3. *Let V be a vector space. If $W \subseteq V$ is a vector space under the vector addition and scalar multiplication operations defined on $V \times V$ and $\mathbb{F} \times V$, respectively, then W is a* ***subspace*** *of V.*

In order for $W \subseteq V$ to be a vector space it must satisfy the statement of Definition 10.1 to the letter, except that the symbol W is substituted for V. Straightaway this means we must have $W \neq \varnothing$ since Axiom VS3 requires that $\mathbf{0} \in W$. Moreover, vector addition must map $W \times W \to W$ and scalar multiplication must map $\mathbb{F} \times W \to W$, which is to say for any $\mathbf{u}, \mathbf{v} \in W$ and $a \in \mathbb{F}$ we must have $\mathbf{u} + \mathbf{v} \in W$ and $a\mathbf{u} \in W$. These observations prove the forward implication in the following theorem.

Theorem 10.4. *Let V be a vector space and $\varnothing \neq W \subseteq V$. Then W is a subspace of V if and only if $a\mathbf{u} \in W$ and $\mathbf{u} + \mathbf{v} \in W$ for all $a \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in W$.*

Proof. We need only prove the reverse implication. So, suppose that for any $a \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in W$, it is true that $a\mathbf{u} \in W$ and $\mathbf{u} + \mathbf{v} \in W$. Then vector addition maps $W \times W \to W$ and scalar multiplication maps $\mathbb{F} \times W \to W$, and it remains to confirm that W satisfies the eight axioms in Definition 10.1. But it is clear that Axioms VS1, VS2, VS5, VS6, VS7, and VS8 must hold. For instance if $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ since $\mathbf{u}, \mathbf{v} \in V$ and V is given to be a vector space, and so Axiom VS1 is confirmed.

Let $\mathbf{u} \in W$. Since $a\mathbf{u} \in W$ for any $a \in \mathbb{F}$, it follows that $(-1)\mathbf{u} \in W$ in particular. Now, $(-1)\mathbf{u} = -\mathbf{u}$ by Proposition 10.2, and so $-\mathbf{u} \in W$. That is, for every $\mathbf{u} \in W$ we find that $-\mathbf{u} \in W$ as well, where $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$. This shows that Axiom VS4 holds for W.

Finally, since $a\mathbf{u} \in W$ for any $a \in \mathbb{F}$, it follows that $0\mathbf{u} \in W$. By Proposition 10.2 we have $0\mathbf{u} = \mathbf{0}$, so $\mathbf{0} \in W$ and Axiom VS3 holds for W.

We conclude that $W \subseteq V$ is a vector space under the vector addition and scalar multiplication operations defined on $V \times V$ and $\mathbb{F} \times V$, respectively. Therefore W is a subspace of V by Definition 10.3. ∎

Definition 10.5. *A vector $\mathbf{v}$ is called a* ***linear combination*** *of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ if there exist scalars $a_1, \ldots, a_n$ such that $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$.*

Definition 10.6. *Let V be a vector space and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. We say $\mathbf{v}_1, \ldots, \mathbf{v}_n$ **span** V, or V is **spanned** by the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, if for every $\mathbf{v} \in V$ there exist scalars $a_1, \ldots, a_n$ such that $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$.*

Thus vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span V if and only if every vector in V is expressible as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. If we define the **span** of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to be the set

$$\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = \left\{ \sum_{i=1}^{n} a_i \mathbf{v}_i : a_1, \ldots, a_n \in \mathbb{F} \right\},$$

then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V if and only if $V = \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. More generally, for any $S \subseteq V$, define $\operatorname{Span}(S)$ to be the set of all linear combinations of vectors in S.

Definition 10.7. *Let V be a vector space and $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$ be nonempty. If the equation*

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0} \tag{10.1}$$

*admits only the trivial solution $c_1 = \cdots = c_n = 0$, then we call S a **linearly independent** set and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ **linearly independent** vectors. Otherwise we call S a **linearly dependent** set and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ **linearly dependent** vectors. A **basis** for a vector space V is a linearly independent set of vectors $\mathcal{B}$ that spans V.*

We say V is a **finite-dimensional** vector space if V possesses a finite basis. The proof of the following proposition is given in §3.6 of [LAN].

Proposition 10.8. *Let V be a vector space such that $V = \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. If $\mathbf{u}_1, \ldots, \mathbf{u}_n \in V$ for some $n > m$, then the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent.*

Theorem 10.9. *If $\mathcal{B}_1$ and $\mathcal{B}_2$ are two bases for a finite-dimensional vector space V, then* $\operatorname{card}(\mathcal{B}_1) = \operatorname{card}(\mathcal{B}_2)$.

Proof. Suppose $\mathcal{B}_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ and $\mathcal{B}_2 = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are two bases for V, so that $\operatorname{card}(\mathcal{B}_1) = m$ and $\operatorname{card}(\mathcal{B}_2) = n$.

Since $\operatorname{Span}(\mathcal{B}_1) = V$, if $n > m$ then $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent by Proposition 10.8, which contradicts the hypothesis that $\mathcal{B}_2$ is a basis for V. Hence $n \leq m$.

Since $\operatorname{Span}(\mathcal{B}_2) = V$, if $n < m$ then $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent by Proposition 10.8, which contradicts the hypothesis that $\mathcal{B}_1$ is a basis for V. Hence $n \geq m$.

Therefore $m = n$, which is to say $\operatorname{card}(\mathcal{B}_1) = \operatorname{card}(\mathcal{B}_2)$. ∎

Since the cardinality of any two bases of a finite-dimensional vector space V is the same, we may meaningfully define the dimension of V as follows.

Definition 10.10. *If V is a finite-dimensional vector space and $\mathcal{B}$ is any basis, then the **dimension** of V is $\dim(V) = \operatorname{card}(\mathcal{B})$.*

Definition 10.11. *Let V be a vector space and $S \subseteq V$ a nonempty set. We call $B \subseteq S$ a **maximal linearly independent subset of S** if the following hold:*

1. *B is a linearly independent set.*

2. *For all $A \subseteq S$ with* $\text{card}(A) > \text{card}(B)$*, A is a linearly dependent set.*

Proposition 10.12. *Let V be a vector space, and let $S \subseteq V$ be a finite set such that $V = \text{Span}(S)$. Then*
1. $\dim(V) \leq \text{card}(S)$.
2. *If $B \subseteq S$ is a maximal linearly independent subset of S, then B is a basis for V.*

Proof.
Proof of Part (1). By Proposition 10.8 any set containing more than $\text{card}(S)$ vectors in V must be linearly dependent, so if $\mathcal{B}$ is any basis for V, then we must have $\dim(V) = \text{card}(\mathcal{B}) \leq \text{card}(S)$.

Proof of Part (2). Suppose that $B \subseteq S$ is a maximal linearly independent subset of S. Reindexing the elements of S if necessary, we may assume that $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. If $r = n$, then $B = S$, and so B spans V and we straightaway conclude that B is a basis for V and we're done. Suppose, then, that $1 \leq r < n$. For each $1 \leq i \leq n - r$ let

$$B_i = B \cup \{\mathbf{v}_{r+i}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+i}\}.$$

The set B_i is linearly dependent since $\text{card}(B_i) > \text{card}(B)$, and so there exist scalars $a_{i1}, \ldots, a_{ir}, b_i$, not all zero, such that

$$a_{i1}\mathbf{v}_1 + \cdots + a_{ir}\mathbf{v}_r + b_i\mathbf{v}_{r+i} = \mathbf{0}. \tag{10.2}$$

We must have $b_i \neq 0$, since otherwise (10.2) becomes

$$a_{i1}\mathbf{v}_1 + \cdots + a_{ir}\mathbf{v}_r = \mathbf{0},$$

whereupon the linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_r$ would imply that $a_{i1} = \cdots = a_{ir} = 0$ and so contradict the established fact that not all the scalars $a_{i1}, \ldots, a_{ir}, b_i$ are zero! From the knowledge that $b_i \neq 0$ we may write (10.2) as

$$\mathbf{v}_{r+i} = -\frac{a_{i1}}{b_i}\mathbf{v}_1 - \cdots - \frac{a_{ir}}{b_i}\mathbf{v}_r = \sum_{j=1}^{r} \frac{a_{ij}}{-b_i}\mathbf{v}_j = \sum_{j=1}^{r} d_{ij}\mathbf{v}_j, \tag{10.3}$$

where we define $d_{ij} = -a_{ij}/b_i$ for each $1 \leq i \leq n - r$ and $1 \leq j \leq r$. Hence the vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are each expressible as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

Let $\mathbf{u} \in V$ be arbitrary. Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span V there exist scalars $c_1, \ldots, c_n$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n,$$

and then from (10.3) we have

$$\begin{aligned}
\mathbf{u} &= c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r + \sum_{i=1}^{n-r} c_{r+i}\mathbf{v}_{r+i} = \sum_{j=1}^{r} c_j\mathbf{v}_j + \sum_{i=1}^{n-r}\left(c_{r+i}\sum_{j=1}^{r} d_{ij}\mathbf{v}_j\right) \\
&= \sum_{j=1}^{r} c_j\mathbf{v}_j + \sum_{i=1}^{n-r}\sum_{j=1}^{r} c_{r+i}d_{ij}\mathbf{v}_j = \sum_{j=1}^{r} c_j\mathbf{v}_j + \sum_{j=1}^{r}\sum_{i=1}^{n-r} c_{r+i}d_{ij}\mathbf{v}_j \\
&= \sum_{j=1}^{r}\left(c_j\mathbf{v}_j + \sum_{i=1}^{n-r} c_{r+i}d_{ij}\mathbf{v}_j\right) = \sum_{j=1}^{r}\left(c_j + \sum_{i=1}^{n-r} c_{r+i}d_{ij}\right)\mathbf{v}_j.
\end{aligned}$$

Setting

$$\hat{c}_j = c_j + \sum_{i=1}^{n-r} c_{r+i} d_{ij}$$

for each $1 \leq j \leq r$, we finally obtain

$$\mathbf{u} = \hat{c}_1 \mathbf{v}_1 + \cdots + \hat{c}_r \mathbf{v}_r$$

and so conclude that $\mathbf{u} \in \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = \operatorname{Span}(B)$.

Therefore $V = \operatorname{Span}(B)$, and so B is a basis for V. ■

A useful concept closely related to that given in Definition 10.11, but not quite the same, is the following.[1]

Definition 10.13. *Let V be a vector space. A set $B \subseteq V$ is a **maximal linearly independent set in V** if the following are true:*

1. *B is a linearly independent set.*
2. *For all $\mathbf{w} \in V$ such that $\mathbf{w} \notin B$, the set $B \cup \{\mathbf{w}\}$ is linearly dependent.*

Proposition 10.14. *If V is a vector space and S a maximal linearly independent set in V, then S is a basis for V.*

Proof. Suppose that V is a vector space and $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a maximal linearly independent set in V. Let $\mathbf{u} \in V$. Then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}\}$ is linearly dependent, and so there exist scalars $a_0, \ldots, a_n$ not all zero such that

$$a_0 \mathbf{u} + a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n = \mathbf{0}. \tag{10.4}$$

Now, if a_0 were 0 we would obtain $a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n = \mathbf{0}$, whereupon the linear independence of S would imply that $a_1 = \cdots = a_n = 0$ and so contradict the established fact that not all the scalars $a_0, \ldots, a_n$ are zero. Hence we must have $a_0 \neq 0$, and (10.4) gives

$$\mathbf{u} = -\frac{a_1}{a_0}\mathbf{v}_1 - \cdots - \frac{a_n}{a_0}\mathbf{v}_n.$$

That is, every vector in V is expressible as a linear combination of vectors in S, so that $\operatorname{Span}(S) = V$ and we conclude that S is a basis for V. ■

Theorem 10.15. *Let V be a finite-dimensional vector space, and let $S \subseteq V$ with* $\operatorname{card}(S) = \dim(V)$.

1. *If S is a linearly independent set, then S is a basis for V.*
2. *If* $\operatorname{Span}(S) = V$*, then S is a basis for V.*

[1] In comparing the two definitions, note the replacement of the words "subset of" with "set in."

Proof.
Proof of Part (1). Setting $n = \dim(V)$, suppose $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$ is a linearly independent set. Any basis for V will span V and have n vectors, so by Proposition 10.8 the set $S \cup \{\mathbf{w}\}$ must be linearly dependent for every $\mathbf{w} \in V$ such that $\mathbf{w} \notin S$. Hence S is a maximal linearly independent set in V, and therefore S is a basis for V by Proposition 10.14.

Proof of Part (2). Again set $n = \dim(V)$, and suppose $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is such that $\mathrm{Span}(S) = V$. Assume S is not a basis for V. Then S must not be a linearly independent set. Let $B \subseteq S$ be a maximal linearly independent subset of S. Then B cannot contain all of the vectors in S, so $\mathrm{card}(B) < \mathrm{card}(S) = n$. By Proposition 10.12(2) it follows that B is a basis for V, and so

$$\dim(V) = \mathrm{card}(B) < n.$$

Since this is a contradiction, we conclude that S must be a linearly independent set and therefore S is a basis for V. ■

Theorem 10.16. *Let V be a vector space with $\dim(V) = n > 0$. If $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$ are linearly independent vectors for some $r < n$, then vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \in V$ may be found such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V.*

Proof. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$ are linearly independent vectors, where $r < n$. The set $S_r = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ cannot be a basis for V since by Definition 10.10 any basis for V must contain n vectors. Hence S_r cannot be a maximal linearly independent set in V by Theorem 10.14, and so there must exist some vector $\mathbf{v}_{r+1} \in V$ such that the set

$$S_{r+1} = S_r \cup \{\mathbf{v}_{r+1}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}\}$$

is linearly independent. Now, if $r + 1 = n$, then Theorem 10.15(1) implies that S_{r+1} is a basis for V and the proof is done. If $r + 1 < n$, then we repeat the arguments made above to obtain successive sets of linearly independent vectors

$$S_{r+i} = S_{r+i-1} \cup \{\mathbf{v}_{r+i}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{r+i}\}$$

until such time that $r + i = n$, at which point the linearly independent set

$$S_n = S_{n-1} \cup \{\mathbf{v}_n\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

will be a basis for V. ■

Theorem 10.17. *Let V be a finite-dimensional vector space, and let W be a subspace of V. Then*

1. *W is finite-dimensional.*
2. $\dim(W) \leq \dim(V)$.
3. *If $\dim(W) = \dim(V)$, then $W = V$.*

Proof. If $W = \{\mathbf{0}\}$, then all three conclusions of the theorem follow trivially. Thus, we will henceforth assume $W \neq \{\mathbf{0}\}$, so that $\dim(V) = n \geq 1$.

Proof of Part (1). Suppose W is infinite-dimensional. Let $\mathbf{w}_1$ be a nonzero vector in W. The set $\{\mathbf{w}_1\}$ cannot be a maximal linearly independent set in W since otherwise Proposition 10.14 would imply that $\{\mathbf{w}_1\}$ is a basis for W and hence $\dim(W) = 1$, a contradiction. Thus for some $k \geq 2$ additional vectors $\mathbf{w}_2, \ldots, \mathbf{w}_k \in W$ may be found such that $S_k = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is a linearly independent set of vectors in W. However, for no $k \in \mathbb{N}$ can S_k be a maximal set of linearly independent vectors in W, since otherwise Proposition 10.14 would imply that $\dim(W) = k$. It follows that there exists, in particular, a linearly independent set

$$\{\mathbf{w}_1, \ldots, \mathbf{w}_{n+1}\} \subseteq W \subseteq V,$$

which is impossible since by Proposition 10.8 there can be no linearly independent set in V containing more than n vectors. Therefore W must be finite-dimensional.

Proof of Part (2). By Part (1) it is known that W is finite-dimensional, so there exists a basis $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ for W, where $m \in \mathbb{N}$. Since $\mathcal{B}$ is a linearly independent set in V, and by Proposition 10.8 there can be no linearly independent set in V containing more than $\dim(V) = n$ vectors, it follows that $\dim(W) = m \leq n = \dim(V)$.

Proof of Part (3). Suppose that $\dim(W) = \dim(V) = n$, where n is some integer since V is given to be finite-dimensional. Let $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis for W, so that $W = \mathrm{Span}(\mathcal{B})$. Since $\dim(V) = n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n \in V$ are linearly independent, $\mathcal{B}$ is a basis for V by Theorem 10.15(1). Thus $V = \mathrm{Span}(\mathcal{B})$, and we have $V = W$. ■

We now narrow our focus. Throughout this chapter our vector space V over $\mathbb{F}$ shall always be euclidean n-space $\mathbb{R}^n$ over the field of real numbers $\mathbb{R}$, and any subspace we consider shall be a subspace of $\mathbb{R}^n$. The elements of $\mathbb{R}^n$ shall be represented by column vectors:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ldots, x_n \in \mathbb{R}^n \right\}.$$

As usual we define vector addition and scalar multiplication in $\mathbb{R}^n$ componentwise:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad a \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ \vdots \\ ax_n \end{bmatrix}$$

A nonempty set $S \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ (and hence may itself be called a vector space) if and only if it is closed under vector addition and scalar multiplication; that is, $\varnothing \neq S \subseteq \mathbb{R}^n$ is a vector space if and only if $\mathbf{x} + \mathbf{y} \in S$ and $a\mathbf{x} \in S$ for any $\mathbf{x}, \mathbf{y} \in S$ and $a \in \mathbb{R}$.

Recall the Kronecker delta function,

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$$

Define the vector $\mathbf{e}_k \in \mathbb{R}^n$ to have ith component $[\mathbf{e}_k]_i = \delta_{ik}$ for each $1 \leq i \leq n$. It is straightforward to show that the set of vectors $\mathcal{E}_n = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$, called the **standard basis**. Therefore

$$\dim(\mathbb{R}^n) = n$$

for any $n \in \mathbb{N}$.

10.2 – Linear Mappings

A **mapping** (or **transformation**) we take to be a function between vector spaces or subsets of vector spaces. Of particular importance is the notion of a linear mapping.

Definition 10.18. *Let V and W be vector spaces over $\mathbb{F}$. A mapping $L : V \to W$ is called a* ***linear mapping*** *if the following properties hold.*

LT1. $L(\mathbf{u}+\mathbf{v}) = L(\mathbf{u}) + L(\mathbf{v})$ *for all* $\mathbf{u}, \mathbf{v} \in V$

LT2. $L(a\mathbf{u}) = aL(\mathbf{u})$ *for all* $a \in \mathbb{F}$ *and* $\mathbf{u} \in V$.

The **zero mapping** $O : V \to W$ given by $O(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$ is a linear mapping. If $U \subseteq V$, then the symbol O_U may sometimes be used to denote the restriction of $O : V \to W$ to U.

A **linear operator** is a linear mapping $L : V \to V$, which may also be referred to as a **linear operator on** V. The **identity operator** I is the linear operator on V given by $I_V(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$.

Given vector spaces V and W over $\mathbb{F}$, the symbol $\mathcal{L}(V, W)$ will be used to denote the set of all linear mappings $V \to W$; that is,

$$\mathcal{L}(V, W) = \{L : V \to W \mid L \text{ is a linear tranformation}\}.$$

We also define $\mathcal{L}(V) = \mathcal{L}(V, V)$; that is, $\mathcal{L}(V)$ denotes the set of all linear operators on V. As is shown in §4.1 of [LAN], the collection $\mathcal{L}(V, W)$ is a vector space under the operations of mapping addition and scalar multiplication given in Definition 10.18.

Definition 10.19. *Let $F : V \to W$ be a mapping. The* ***image*** *of F is the set*

$$\operatorname{Img}(F) = \{\mathbf{w} \in W : F(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\},$$

and the ***null space*** *of F is the set*

$$\operatorname{Nul}(F) = \{\mathbf{v} \in V : F(\mathbf{v}) = \mathbf{0}\}.$$

Note that the image of F is the same as the range of F. As with any function, the symbol $\operatorname{Dom}(F)$ denotes the domain of a mapping F.

Proposition 10.20. *If $L \in \mathcal{L}(V, W)$, then the following hold.*

1. $L(\mathbf{0}) = \mathbf{0}$
2. $L(-\mathbf{v}) = -L(\mathbf{v})$ *for any* $\mathbf{v} \in V$.
3. *For any* $a_1, \dots, a_n \in \mathbb{F}$, $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$,
$$L\left(\sum_{k=1}^{n} a_k \mathbf{v}_k\right) = \sum_{k=1}^{n} a_k L(\mathbf{v}_k).$$
4. $\operatorname{Img}(L)$ *is a subspace of* W.
5. $\operatorname{Nul}(L)$ *is a subspace of* V.
6. *If* $\operatorname{Nul}(L) = \{\mathbf{0}\}$ *and* $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ *are linearly independent, then* $L(\mathbf{v}_1), \dots, L(\mathbf{v}_n) \in W$ *are linearly independent.*

Proof.
Proof of Part (1). Using the linearity property LT1, we have

$$L(\mathbf{0}) = L(\mathbf{0} + \mathbf{0}) = L(\mathbf{0}) + L(\mathbf{0}).$$

Subtracting $L(\mathbf{0})$ from the leftmost and rightmost sides then gives

$$L(\mathbf{0}) - L(\mathbf{0}) = [L(\mathbf{0}) + L(\mathbf{0})] - L(\mathbf{0}),$$

and thus $\mathbf{0} = L(\mathbf{0})$.

Proof of Part (2). Let $\mathbf{v} \in V$ be arbitrary. Using property LT1 and part (1), we have

$$L(\mathbf{v}) + L(-\mathbf{v}) = L(\mathbf{v} + (-\mathbf{v})) = L(\mathbf{0}) = \mathbf{0}.$$

This shows that $L(-\mathbf{v})$ is the additive inverse of $L(\mathbf{v})$. That is, $L(-\mathbf{v}) = -L(\mathbf{v})$.

Proof of Part (3). We have $L(a_1\mathbf{v}_1) = a_1 L(\mathbf{v}_1)$ by property LT2. Let $n \in \mathbb{N}$ and suppose that

$$(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1 L(\mathbf{v}_1) + \cdots + a_n L(\mathbf{v}_n) \tag{10.5}$$

for any $a_1, \ldots, a_n \in \mathbb{F}$, $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. Let $a_1, \ldots, a_{n+1} \in \mathbb{F}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in V$ be arbitrary. Then

$$\begin{aligned} L\Big(\sum\nolimits_{i=1}^{n+1} a_i\mathbf{v}_i\Big) &= L\big((a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + a_{n+1}\mathbf{v}_{n+1}\big) \\ &= L(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + L(a_{n+1}\mathbf{v}_{n+1}) && \text{Property LT1} \\ &= a_1 L(\mathbf{v}_1) + \cdots + a_n L(\mathbf{v}_n) + L(a_{n+1}\mathbf{v}_{n+1}) && \text{Hypothesis (10.5)} \\ &= a_1 L(\mathbf{v}_1) + \cdots + a_n L(\mathbf{v}_n) + a_{n+1} L(\mathbf{v}_{n+1}) && \text{Property LT2} \\ &= \sum\nolimits_{i=1}^{n+1} a_i L(\mathbf{v}_i) \end{aligned}$$

The proof is complete by the Principle of Induction.

Proof of Part (4). We have $L(\mathbf{0}) = \mathbf{0}$ from part (1), and so $\mathbf{0} \in \text{Img}(L)$.

Suppose that $\mathbf{w}_1, \mathbf{w}_2 \in \text{Img}(L)$. Then there exist vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $L(\mathbf{v}_1) = \mathbf{w}_1$ and $L(\mathbf{v}_2) = \mathbf{w}_2$. Now, since $\mathbf{v}_1 + \mathbf{v}_2 \in V$ and

$$L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2,$$

we conclude that $\mathbf{w}_1 + \mathbf{w}_2 \in \text{Img}(L)$. Hence $\text{Img}(L)$ is closed under vector addition.

Finally, let $a \in \mathbb{R}$ and suppose $\mathbf{w} \in \text{Img}(L)$. Then there exists some $\mathbf{v} \in V$ such that $L(\mathbf{v}) = \mathbf{w}$, and since $a\mathbf{v} \in V$ and

$$L(a\mathbf{v}) = aL(\mathbf{v}) = a\mathbf{w},$$

we conclude that $a\mathbf{w} \in \text{Img}(L)$. Hence $\text{Img}(L)$ is closed under scalar multiplication.

Therefore $\text{Img}(L) \subseteq W$ is a subspace by Theorem 10.4.

Proof of Part (5). Since $L(\mathbf{0}) = \mathbf{0}$ we immediately obtain $\mathbf{0} \in \text{Nul}(L)$.

Suppose that $\mathbf{v}_1, \mathbf{v}_2 \in \mathrm{Nul}(L)$. Then $L(\mathbf{v}_1) = L(\mathbf{v}_2) = \mathbf{0}$, and since

$$L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

it follows that $\mathbf{v}_1 + \mathbf{v}_2 \in \mathrm{Nul}(L)$ and so $\mathrm{Nul}(L)$ is closed under vector addition.

Finally, let $a \in \mathbb{R}$ and suppose $\mathbf{v} \in \mathrm{Nul}(L)$. Then $L(\mathbf{v}) = \mathbf{0}$, and since

$$L(a\mathbf{v}) = aL(\mathbf{v}) = a\mathbf{0} = \mathbf{0}$$

by Proposition 10.2, we conclude that $a\mathbf{v} \in \mathrm{Nul}(L)$ and so $\mathrm{Nul}(L)$ is closed under scalar multiplication.

Therefore $\mathrm{Nul}(L) \subseteq V$ is a subspace by Theorem 10.4.

Proof of Part (6). Suppose $\mathrm{Nul}(L) = \{\mathbf{0}\}$ and $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ are linearly independent. Let $a_1, \dots, a_n \in \mathbb{F}$ be such that

$$a_1 L(\mathbf{v}_1) + \cdots + a_n L(\mathbf{v}_n) = \mathbf{0}.$$

From this we obtain

$$L(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = \mathbf{0},$$

and since $\mathrm{Nul}(L) = \{\mathbf{0}\}$ it follows that

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}.$$

Now, since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, it follows that $a_1 = \cdots = a_n = 0$. Therefore the vectors $L(\mathbf{v}_1), \dots, L(\mathbf{v}_n)$ in W are linearly independent. ■

Proposition 10.21. *Let V be a finite-dimensional vector space over $\mathbb{F}$, and let $L \in \mathcal{L}(V)$. Then the following are equivalent.*

1. *L is injective.*
2. *L is surjective.*
3. $\mathrm{Nul}(L) = \{\mathbf{0}\}$.

Proof.
(1) ↔ (3). Suppose that $L \in \mathcal{L}(V)$ is injective. Let $\mathbf{v} \in \mathrm{Nul}(L)$, so that $L(\mathbf{v}) = \mathbf{0}$. By Proposition 10.20 we have $L(\mathbf{0}) = \mathbf{0}$ also, and since L is injective it follows that $\mathbf{v} = \mathbf{0}$. Hence $\mathrm{Nul}(L) \subseteq \{\mathbf{0}\}$, and $L(\mathbf{0}) = \mathbf{0}$ shows that $\{\mathbf{0}\} \subseteq \mathrm{Nul}(L)$. Therefore $\mathrm{Nul}(L) = \{\mathbf{0}\}$.

Next, suppose that $\mathrm{Nul}(L) = \{\mathbf{0}\}$. Suppose that $L(\mathbf{v}_1) = L(\mathbf{v}_2)$, so $L(\mathbf{v}_1) - L(\mathbf{v}_2) = \mathbf{0}$. Then

$$L(\mathbf{v}_1 - \mathbf{v}_2) = L(\mathbf{v}_1) - L(\mathbf{v}_2) = \mathbf{0}$$

shows that $\mathbf{v}_1 - \mathbf{v}_2 \in \mathrm{Nul}(L) = \{\mathbf{0}\}$ and thus $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$. Therefore $\mathbf{v}_1 = \mathbf{v}_2$ and we conclude that L is injective.

(2) ↔ (3). Suppose L is surjective. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V, so $\dim(V) = n$. For any $\mathbf{v} \in V$ there exists some $\mathbf{u} \in V$ such that $L(\mathbf{u}) = \mathbf{v}$, and since $\mathbf{u} = \sum_{k=1}^n a_k\mathbf{v}_k$ for some $a_1, \dots, a_n \in \mathbb{F}$, by Proposition 10.20(3) we have $\mathbf{v} = \sum_{k=1}^n a_k L(\mathbf{v}_k)$, which shows that the set $S = \{L(\mathbf{v}_1), \dots, L(\mathbf{v}_n)\}$ is such that $\mathrm{Span}(S) = V$. Clearly $\mathrm{card}(S) \le n = \dim(V)$, but also

$\text{card}(S) \geq \dim(V)$ by Proposition 10.12, and therefore $\text{card}(S) = \dim(V) = n$. By Theorem 10.15(2) it follows that S is a basis for V, so S is a linearly independent set. Now,

$$L(\mathbf{u}) = \mathbf{0} \;\Rightarrow\; \sum_{k=1}^{n} a_k L(\mathbf{v}_k) = \mathbf{0} \;\Rightarrow\; a_1 = \cdots = a_n = 0 \;\Rightarrow\; \mathbf{u} = \mathbf{0},$$

and therefore $\text{Nul}(L) = \{\mathbf{0}\}$.

Now suppose that $\text{Nul}(L) = \{\mathbf{0}\}$. By Proposition 10.20(6) the set S is linearly independent. Now, $\text{Img}(L)$ is a subspace of V by Proposition 10.20(4), and since $\text{Span}(S) = \text{Img}(L)$ it follows that S is a basis for $\text{Img}(L)$. Thus $\dim(\text{Img}(L)) = \text{card}(S) = n = \dim(V)$, and by Theorem 10.17(3) we conclude that $\text{Img}(L) = V$. That is, L is surjective. ■

Definition 10.22. *Given mappings $F : U \to V$ and $G : V \to W$, the* ***composition*** *of G with F is the mapping $G \circ F : U \to W$ given by*

$$(G \circ F)(\mathbf{v}) = G(F(\mathbf{v}))$$

for all $\mathbf{v} \in U$.

Proposition 10.23. *Let V_1, V_2, V_3 be vector spaces over $\mathbb{F}$. If $L_1 : V_1 \to V_2$ and $L_2 : V_2 \to V_3$ are linear mappings, then the composition $L_2 \circ L_1 : V_1 \to V_3$ is linear.*

Proof. For any $\mathbf{u}, \mathbf{v} \in V_1$ we have

$$\begin{aligned}(L_2 \circ L_1)(\mathbf{u} + \mathbf{v}) &= L_2(L_1(\mathbf{u} + \mathbf{v})) = L_2(L_1(\mathbf{u}) + L_1(\mathbf{v})) \\ &= L_2(L_1(\mathbf{u})) + L_2(L_1(\mathbf{v})) = (L_2 \circ L_1)(\mathbf{u}) + (L_2 \circ L_1)(\mathbf{v}),\end{aligned}$$

and for any $a \in \mathbb{F}$ and $\mathbf{u} \in V_1$ we have

$$(L_2 \circ L_1)(a\mathbf{u}) = L_2(L_1(a\mathbf{u})) = L_2(aL_1(\mathbf{u})) = aL_2(L_1(\mathbf{u})) = a(L_2 \circ L_1)(\mathbf{u}).$$

Therefore $L_2 \circ L_1$ is linear. ■

Definition 10.24. *Let $F : V \to W$ be a mapping. We say F is* ***invertible*** *if there exists a mapping $G : W \to V$ such that $G \circ F = I_V$ and $F \circ G = I_W$, in which case G is called the* ***inverse*** *of F and we write $G = F^{-1}$.*

Proposition 10.25. *If $F : V \to W$ is an invertible mapping, then*

$$\text{Img}(F^{-1}) = \text{Dom}(F) = V \quad \textit{and} \quad \text{Dom}(F^{-1}) = \text{Img}(F) = W,$$

and for all $\mathbf{v} \in V$, $\mathbf{w} \in W$,

$$F(\mathbf{v}) = \mathbf{w} \;\Leftrightarrow\; F^{-1}(\mathbf{w}) = \mathbf{v}.$$

Proof. Suppose that $F : V \to W$ is invertible, so that there is a mapping $F^{-1} : W \to V$ such that $F^{-1} \circ F = I_V$ and $F \circ F^{-1} = I_W$. From this it follows immediately that $\text{Img}(F^{-1}) \subseteq V$ and $\text{Img}(F) \subseteq W$.

Let $\mathbf{v} \in V$, so that $F(\mathbf{v}) = \mathbf{w}$ for some $\mathbf{w} \in W$. Then

$$F^{-1}(\mathbf{w}) = F^{-1}(F(\mathbf{v})) = (F^{-1} \circ F)(\mathbf{v}) = I_V(\mathbf{v}) = \mathbf{v}$$

shows that $\mathbf{v} \in \operatorname{Img}(F^{-1})$, and so $\operatorname{Img}(F^{-1}) = V$ and

$$F(\mathbf{v}) = \mathbf{w} \quad \Rightarrow \quad F^{-1}(\mathbf{w}) = \mathbf{v}$$

for all $\mathbf{v} \in V$.

Next, for any $\mathbf{w} \in W$ we have $F^{-1}(\mathbf{w}) = \mathbf{v}$ for some $\mathbf{v} \in V$, whence

$$F(\mathbf{v}) = F(F^{-1}(\mathbf{w})) = (F \circ F^{-1})(\mathbf{w}) = I_W(\mathbf{w}) = \mathbf{w}$$

shows that $\mathbf{w} \in \operatorname{Img}(F)$, and so $\operatorname{Img}(F) = W$ and

$$F^{-1}(\mathbf{w}) = \mathbf{v} \quad \Rightarrow \quad F(\mathbf{v}) = y$$

for all $\mathbf{w} \in W$. ■

An consequence of the foregoing is the following theorem, the proof of which is left as a straightforward exercise.

Theorem 10.26. *A mapping $F : U \to V$ is invertible if and only if it is a bijection.*

Proposition 10.27. *If $F : U \to V$ and $G : V \to W$ are invertible mappings, then*

$$(G \circ F)^{-1} = F^{-1} \circ G^{-1}.$$

Proof. Suppose $F : U \to V$ and $G : V \to W$ are invertible mappings. Then F and G are bijective by Theorem 10.26, from which it follows that $G \circ F$ is likewise bijective and so $(G \circ F)^{-1} : W \to U$ exists. That is, $G \circ F$ is invertible.

Let $\mathbf{w} \in W$. Then $(G \circ F)^{-1}(\mathbf{w}) = \mathbf{v}$ for some $\mathbf{v} \in U$, and by repeated use of Proposition 10.25 we obtain

$$\begin{aligned} (G \circ F)^{-1}(\mathbf{w}) = \mathbf{v} \quad &\Leftrightarrow \quad (G \circ F)(\mathbf{v}) = \mathbf{w} \quad \Leftrightarrow \quad G(F(\mathbf{v})) = \mathbf{w} \\ &\Leftrightarrow \quad F(\mathbf{v}) = G^{-1}(\mathbf{w}) \quad \Leftrightarrow \quad \mathbf{v} = F^{-1}(G^{-1}(\mathbf{w})). \\ &\Leftrightarrow \quad (F^{-1} \circ G^{-1})(\mathbf{w}) = \mathbf{v} \end{aligned}$$

Hence

$$(G \circ F)^{-1}(\mathbf{w}) = (F^{-1} \circ G^{-1})(\mathbf{w})$$

for all $\mathbf{w} \in W$, and we conclude that $(G \circ F)^{-1} = F^{-1} \circ G^{-1}$. ■

Proposition 10.28. *If $L : V \to W$ is an invertible linear mapping, then its inverse $L^{-1} : W \to V$ is also linear.*

Proof. Suppose that $L : V \to W$ is an invertible linear mapping, and let $L^{-1} : W \to V$ be its inverse. Let $\mathbf{w}_1, \mathbf{w}_2 \in W$. Then $L^{-1}(\mathbf{w}_1)$ and $L^{-1}(\mathbf{w}_2)$ are vectors in V, and by the linearity of L we obtain

$$\begin{aligned} L(L^{-1}(\mathbf{w}_1) + L^{-1}(\mathbf{w}_2)) &= L(L^{-1}(\mathbf{w}_1)) + L(L^{-1}(\mathbf{w}_2)) \\ &= (L \circ L^{-1})(\mathbf{w}_1) + (L \circ L^{-1})(\mathbf{w}_2) \\ &= I_W(\mathbf{w}_1) + I_W(\mathbf{w}_2) = \mathbf{w}_1 + \mathbf{w}_2 \end{aligned}$$

Now,

$$L^{-1}(\mathbf{w}_1+\mathbf{w}_2) = L^{-1}(L(L^{-1}(\mathbf{w}_1)+L^{-1}(\mathbf{w}_2))) = (L^{-1}\circ L)(L^{-1}(\mathbf{w}_1)+L^{-1}(\mathbf{w}_2))$$
$$= I_V(L^{-1}(\mathbf{w}_1)+L^{-1}(\mathbf{w}_2)) = L^{-1}(\mathbf{w}_1)+L^{-1}(\mathbf{w}_2).$$

Next, let $\mathbf{w}\in W$ and let a be a scalar. Then $aL^{-1}(\mathbf{w})$ is a vector in V, and from

$$L(aL^{-1}(\mathbf{w})) = aL(L^{-1}(\mathbf{w})) = a(L\circ L^{-1})(\mathbf{w}) = aI_W(\mathbf{w}) = a\mathbf{w}$$

we obtain

$$L^{-1}(a\mathbf{w}) = L^{-1}(L(aL^{-1}(\mathbf{w}))) = (L^{-1}\circ L)(aL^{-1}(\mathbf{w})) = I_V(aL^{-1}(\mathbf{w})) = aL^{-1}(\mathbf{w}).$$

Therefore L^{-1} is a linear mappings. ■

Definition 10.29. *Given linear mappings $L_1, L_2 : V \to W$, we define the mapping $L_1+L_2 : V \to W$ by*

$$(L_1+L_2)(\mathbf{v}) = L_1(\mathbf{v}) + L_2(\mathbf{v})$$

for each $\mathbf{v}\in V$.

Given linear mapping $L : V \to W$ and $a\in\mathbb{F}$, we define $aL : V\to W$ by

$$(aL)(\mathbf{v}) = aL(\mathbf{v})$$

for each $\mathbf{v}\in V$. In particular we define $-L = (-1)L$.

Proposition 10.30. *Let $F_1, F_2 : U\to V$ and $G_1, G_2 : V\to W$ be mappings, and let $a\in\mathbb{F}$. Then*

1. $(G_1\pm G_2)\circ F_1 = G_1\circ F_1 \pm G_2\circ F_1$
2. $G_1\circ(F_1\pm F_2) = G_1\circ F_1\pm G_1\circ F_2$ *if G_1 is linear and V and W are vector spaces.*
3. $(aG_1)\circ F_1 = a(G_1\circ F_1)$
4. $G_1\circ(aF_1) = a(G_1\circ F_1)$ *if G_1 is linear and V and W are vector spaces.*

Proof.

Proof of Part (1). For any $\mathbf{u}\in U$

$$((G_1+G_2)\circ F_1)(\mathbf{u}) = (G_1+G_2)(F_1(\mathbf{u})) = G_1(F_1(\mathbf{u})) + G_2(F_1(\mathbf{u}))$$
$$= (G_1\circ F_1)(\mathbf{u}) + (G_2\circ F_1)(\mathbf{u}) = (G_1\circ F_1 + G_2\circ F_1)(\mathbf{u}),$$

and therefore $(G_1+G_2)\circ F_1 = G_1\circ F_1 + G_2\circ F_1$. The proof that $(G_1-G_2)\circ F_1 = G_1\circ F_1 - G_2\circ F_1$ is similar.

Proof of Part (2). For any $\mathbf{u}\in U$

$$(G_1\circ(F_1+F_2))(\mathbf{u}) = G_1((F_1+F_2)(\mathbf{u})) = G_1(F_1(\mathbf{u})) + F_2(\mathbf{u}))$$
$$= G_1(F_1(\mathbf{u})) + G_1(F_2(\mathbf{u})) = (G_1\circ F_1)(\mathbf{u}) + (G_1\circ F_2)(\mathbf{u})$$
$$= (G_1\circ F_1 + G_1\circ F_2)(\mathbf{u}),$$

where the third equality obtains from the linearity of G_1. Therefore

$$G_1 \circ (F_1 + F_2) = G_1 \circ F_1 + G_1 \circ F_2$$

if G_1 is linear. The proof that $G_1 \circ (F_1 - F_2) = G_1 \circ F_1 - G_1 \circ F_2$ if G_1 is linear is similar.

Proof of Part (3). For any $\mathbf{u} \in U$

$$((aG_1) \circ F_1)(\mathbf{u}) = (aG_1)(F_1(\mathbf{u})) = aG_1(F_1(\mathbf{u})) = a(G_1 \circ F_1)(\mathbf{u}),$$

and therefore $(aG_1) \circ F_1 = a(G_1 \circ F_1)$.

Proof of Part (4). Suppose that G_1 is a linear mapping. For any $\mathbf{u} \in U$

$$(G_1 \circ (aF_1))(\mathbf{u}) = G_1((aF_1)(\mathbf{u})) = G_1(aF_1(\mathbf{u})) = aG_1(F_1(\mathbf{u})) = a(G_1 \circ F_1)(\mathbf{u}),$$

where the third equality obtains from the linearity of G_1. Therefore $G_1 \circ (aF_1) = a(G_1 \circ F_1)$ if G_1 is linear. ■

For convenience we present the following theorem which puts together many salient results concerning linear operators.

Theorem 10.31 (Invertible Operator Theorem). *Let V be a finite-dimensional vector space, and suppose $L \in \mathcal{L}(V)$. Then the following statements are equivalent.*

1. *L is invertible.*
2. *L is an isomorphism.*
3. *L is injective.*
4. *L is surjective.*
5. $\mathrm{Nul}(L) = \{\mathbf{0}\}$.
6. $\mathrm{rank}(L) = \dim(V)$.

Proof.
(1) ⇒ (2): If L is invertible, then it is a bijection by Theorem 10.26, and hence an isomorphism.

(2) ⇒ (3) ⇒ (4) ⇒ (5): If L is an isomorphism, then it is immediate that L is injective. Statements (3), (4), and (5) are equivalent by Proposition 10.21.

(5) ⇒ (6): Suppose $\mathrm{Nul}(L) = \{\mathbf{0}\}$. Then L is surjective by Proposition 10.21, which is to say $\mathrm{Img}(L) = V$, and therefore $\mathrm{rank}(L) = \dim(V)$.

(6) ⇒ (1): Suppose $\mathrm{rank}(L) = \dim(V)$. Then $\dim(\mathrm{Img}(L)) = \dim(V)$, and since $\mathrm{Img}(L)$ is a subspace of V by Proposition 10.20(4), by Theorem 10.17(3) it follows that $\mathrm{Img}(L) = V$. Hence L is surjective, whereupon Proposition 10.21 gives that L is also injective, and then Theorem 10.26 implies that L is invertible. ■

Let V and W be finite-dimensional vector spaces over $\mathbb{R}$ with bases $\mathcal{B} = (\mathbf{b}_j)_{j=1}^n$ and $\mathcal{C} = (\mathbf{c}_i)_{i=1}^m$, respectively. Given $\mathbf{v} \in V$, there exist $v_1, \dots, v_n \in \mathbb{R}$ such that

$$\mathbf{v} = \sum_{j=1}^n v_j \mathbf{b}_j,$$

and then the **$\mathcal{B}$-coordinates** of $\mathbf{v}$ are

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Let $L \in \mathcal{L}(V, W)$. For each $1 \le j \le n$ we have $L(\mathbf{b}_j) \in W$, so there exist $a_{1j}, \dots, a_{mj} \in \mathbb{R}$ such that

$$L(\mathbf{b}_j) = \sum_{i=1}^m a_{ij} \mathbf{c}_i,$$

and thus the $\mathcal{C}$-coordinates of $L(\mathbf{b}_j)$ are

$$[L(\mathbf{b}_j)]_{\mathcal{C}} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}. \tag{10.6}$$

Now, by the linearity properties of L,

$$L(\mathbf{v}) = \sum_{j=1}^n v_j L(\mathbf{b}_j) = \sum_{j=1}^n v_j \left(\sum_{i=1}^m a_{ij} \mathbf{c}_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n v_j a_{ij} \right) \mathbf{c}_i. \tag{10.7}$$

Thus the $\mathcal{C}$-coordinates of $L(\mathbf{v})$ are

$$[L(\mathbf{v})]_{\mathcal{C}} = \begin{bmatrix} \sum_{j=1}^n v_j a_{1j} \\ \vdots \\ \sum_{j=1}^n v_j a_{mj} \end{bmatrix}$$

Defining the matrix

$$\mathbf{A} = [a_{ij}]_{m,n} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},$$

it's straightforward to check that $\mathbf{A}[\mathbf{v}]_{\mathcal{B}} = [L(\mathbf{v})]_{\mathcal{C}}$. The matrix $\mathbf{A}$ is called the **matrix corresponding to L with respect to $\mathcal{B}$ and $\mathcal{C}$**, or simply the **$\mathcal{BC}$-matrix** of L, and is denoted by $[L]_{\mathcal{BC}}$. Thus

$$[L]_{\mathcal{BC}}[\mathbf{v}]_{\mathcal{B}} = [L(\mathbf{v})]_{\mathcal{C}}.$$

Recalling (10.6), we see that $[L]_{\mathcal{BC}}$ may be defined in terms of its column vectors as

$$[L]_{\mathcal{BC}} = \Big[\; [L(\mathbf{b}_1)]_{\mathcal{C}} \; \cdots \; [L(\mathbf{b}_n)]_{\mathcal{C}} \; \Big]. \tag{10.8}$$

This matrix may be denoted by $[L]$ if context makes clear what bases are under consideration for the vector spaces involved.

Now, if $L \in \mathcal{L}(V, W)$ is given to have $\mathcal{BC}$-matrix $\mathbf{A} = [a_{ij}]_{m,n}$, it is immediate that for any $\mathbf{v} \in V$ with $\mathcal{B}$-coordinates as before we must have $L(\mathbf{v})$ given by (10.7). Therefore there is a bijective correspondence between the elements of $\mathcal{L}(V, W)$ and the elements of $\text{Mat}_{m,n}(\mathbb{R})$.

Now let us return once again to euclidean vector spaces. Let $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Let $\mathcal{E}_n = (\mathbf{e}_j)_{j=1}^n$ and $\mathcal{E}_m = (\boldsymbol{\epsilon}_i)_{i=1}^m$ be the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. For each $1 \leq j \leq n$ we have

$$L(\mathbf{e}_j) = \sum_{i=1}^m a_{ij}\boldsymbol{\epsilon}_i = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = [L(\mathbf{e}_j)]_{\mathcal{E}_m}$$

for some $a_{1j}, \ldots, a_{mj} \in \mathbb{R}$, and thus by (10.8) we have

$$[L]_{\mathcal{E}_n\mathcal{E}_m} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

as the $\mathcal{E}_n\mathcal{E}_m$**-matrix** of L.

As shown in §4.5 of [LAN], a change of bases for V and W will result in a change in the corresponding matrix for $L \in \mathcal{L}(V, W)$. In the case when $L \in \mathcal{L}(V)$, regardless of what basis $\mathcal{B}$ is chosen for V, the matrix $[L]_{\mathcal{B}} := [L]_{\mathcal{BB}}$ will be a square matrix, and as shown in §5.4 of [LAN] the value of the determinant of $[L]_{\mathcal{B}}$ will remain the same. Thus we may define the **determinant** of any $L \in \mathcal{L}(V)$ to be

$$\det(L) = \det([L])$$

without ambiguity.

Remark. Let V be a finite-dimensional vector space. As can be seen from the Invertible Operator Theorem above, in conjunction with the Invertible Matrix Theorem in §5.3 of [LAN], a linear operator $L \in \mathcal{L}(V)$ is invertible iff $[L]$ is invertible iff $\det([L]) \neq 0$ iff $\det(L) \neq 0$. This fact can be used to cast certain upcoming results in the language of matrices and determinants that is frequently more convenient in practical applications.

10.3 – The Norm of a Linear Mapping

The **euclidean norm** for $\mathbb{R}^n$, denoted by $\|\cdot\|$, is defined thus: if $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$, then

$$\|\mathbf{x}\| = \sqrt{\sum_{k=1}^{n} x_k^2}.$$

Henceforth we assume that any euclidean space $\mathbb{R}^n$ is given the euclidean norm, which induces the **euclidean metric** $d_{\mathrm{e}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$d_{\mathrm{e}}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

so that $\mathbb{R}^n$ is taken to be the metric space $(\mathbb{R}^n, d_{\mathrm{e}})$. It is straightforward to check that

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\| \quad \text{and} \quad \|a\mathbf{x}\| = |a|\|\mathbf{x}\| \tag{10.9}$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

Presently the definition of the norm of a linear mapping between two euclidean spaces will be given, although the notion may easily be generalized to apply to arbitrary normed vector spaces. First, we define the **open standard unit ball** in $\mathbb{R}^n$ to be

$$\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\},$$

and we define the **closed standard unit ball** in $\mathbb{R}^n$ to be

$$\overline{\mathbb{B}}^n = \mathbb{B}^n \cup \partial\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \le 1\}.$$

We now give the definition of the norm of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$, which like the euclidean norm of a vector $\mathbf{x} \in \mathbb{R}^n$ will be indicated by the symbol $\|\cdot\|$. This will not give rise to any ambiguity.

Definition 10.32. *The **norm** of $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is*

$$\|L\| = \sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} \|L(\mathbf{x})\|.$$

Theorem 10.33.

1. *If $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $\|L\| < \infty$ and L is uniformly continuous on $\mathbb{R}^n$.*
2. *If $L, L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $a \in \mathbb{R}$, then*

$$\|L_1 + L_2\| \le \|L_1\| + \|L_2\| \quad \text{and} \quad \|aL\| = |a|\|L\|.$$

3. *For all $L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ define*

$$d_{\sup}(L_1, L_2) = \|L_1 - L_2\|.$$

   Then $\big(\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), d_{\sup}\big)$ is a metric space.
4. *If $L_1 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $L_2 \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^k)$, then*

$$\|L_2 \circ L_1\| \le \|L_2\|\|L_1\|.$$

Proof.
Proof of Part (1). Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis for $\mathbb{R}^n$. Fix $\mathbf{x} \in \overline{\mathbb{B}}^n$, so $\mathbf{x} = \sum_{k=1}^n x_k \mathbf{e}_k$ with

$$\sqrt{\sum\nolimits_{k=1}^n x_k^2} \leq 1,$$

and thus $|x_k| \leq 1$ for all $1 \leq k \leq n$. Now, recalling Proposition 10.20(3) and also (10.9),

$$\|L(\mathbf{x})\| = \left\| \sum_{k=1}^n x_k L(\mathbf{e}_k) \right\| \leq \sum_{k=1}^n \|x_k L(\mathbf{e}_k)\| = \sum_{k=1}^n |x_k| \|L(\mathbf{e}_k)\| \leq \sum_{k=1}^n \|L(\mathbf{e}_k)\|.$$

We see that $\sum_{k=1}^n \|L(\mathbf{e}_k)\|$ is an upper bound on $\|L(\mathbf{x})\|$ for $\mathbf{x} \in \overline{\mathbb{B}}^n$, and therefore

$$\|L\| = \sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} \|L(\mathbf{x})\| \leq \|L(\mathbf{x})\| < \infty.$$

Proof of Part (2). Suppose $L, L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $a \in \mathbb{R}$. Recalling (10.9), for any $\mathbf{x} \in \overline{\mathbb{B}}^n$ we have

$$\|(L_1 + L_2)(\mathbf{x})\| = \|L_1(\mathbf{x}) + L_2(\mathbf{x})\| \leq \|L_1(\mathbf{x})\| + \|L_2(\mathbf{x})\| = \|L_1\| + \|L_2\|,$$

which shows that $\|L_1\| + \|L_2\|$ is an upper bound for the set

$$S = \{\|(L_1 + L_2)(\mathbf{x})\| : \mathbf{x} \in \overline{\mathbb{B}}^n\},$$

and so the least upper bound for S is at most $\|L_1\| + \|L_2\|$:

$$\|L_1 + L_2\| = \sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} \|(L_1 + L_2)(\mathbf{x})\| = \sup(S) \leq \|L_1\| + \|L_2\|.$$

Next, for any $\mathbf{x} \in \overline{\mathbb{B}}^n$,

$$\|(aL)(\mathbf{x})\| = \|aL(\mathbf{x})\| = |a|\|L(\mathbf{x})\| \leq |a|\|L\|,$$

and so

$$\|aL\| = \sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} \|(aL)(\mathbf{x})\| \leq |a|\|L\|.$$

Proof of Part (3). Suppose $L, L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Clearly $d_{\sup}(L, L) = 0$, $d_{\sup}(L_1, L_2) \geq 0$, and $d_{\sup}(L_1, L_2) = d_{\sup}(L_2, L_1)$. Moreover, by part (2),

$$\begin{aligned} d_{\sup}(L_1, L_2) &= \|L_1 - L_2\| = \|(L_1 - L) + (L - L_2)\| \\ &\leq \|L_1 - L\| + \|L - L_2\| = d_{\sup}(L_1, L) + d_{\sup}(L, L_2). \end{aligned}$$

Finally, suppose that $d_{\sup}(L_1, L_2) = 0$, so that

$$\sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} d_{\mathrm{e}}\big(L_1(\mathbf{x}), L_2(\mathbf{x})\big) = \sup_{\mathbf{x} \in \overline{\mathbb{B}}^n} \|L_1(\mathbf{x}) - L_2(\mathbf{x})\| = 0.$$

Thus $d_{\mathrm{e}}(L_1(\mathbf{x}), L_2(\mathbf{x})) = 0$ for all $\mathbf{x} \in \overline{\mathbb{B}}^n$, and since $(\mathbb{R}^m, d_{\mathrm{e}})$ is a metric space it follows that $L_1(\mathbf{x}) = L_2(\mathbf{x})$ for all $\mathbf{x} \in \overline{\mathbb{B}}^n$. Now let $\mathbf{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ be arbitrary, and let $\mathbf{x} = \mathbf{y}/\|\mathbf{y}\|$. Then $L_1(\mathbf{x}) = L_2(\mathbf{x})$ since $\mathbf{x} \in \overline{\mathbb{B}}^n$. Now,

$$L_1(\mathbf{x}) = L_2(\mathbf{x}) \quad\Rightarrow\quad \|\mathbf{y}\|L_1(\mathbf{x}) = \|\mathbf{y}\|L_2(\mathbf{x}) \quad\Rightarrow\quad L_1(\|\mathbf{y}\|\mathbf{x}) = L_2(\|\mathbf{y}\|\mathbf{x}),$$

giving $L_1(\mathbf{y}) = L_2(\mathbf{y})$ and hence $L_1 = L_2$. Therefore $d_{\sup}$ is a metric for $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Proof of Part (4). Suppose $L_1 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $L_2 \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^k)$. We have

$$\|L_2\| = \sup_{\mathbf{x}\in\overline{\mathbb{B}}^m} \|L_2(\mathbf{x})\| \quad \text{and} \quad \|L_1\| = \sup_{\mathbf{x}\in\overline{\mathbb{B}}^n} \|L_1(\mathbf{x})\|.$$

Fix $\mathbf{x} \in \overline{\mathbb{B}}^n$, and first suppose that $L_1(\mathbf{x}) \neq \mathbf{0}$. Then

$$\frac{1}{\|L_1(\mathbf{x})\|}\|L_2(L_1(\mathbf{x}))\| = \left\|\frac{1}{\|L_1(\mathbf{x})\|}L_2(L_1(\mathbf{x}))\right\| = \left\|L_2\left(\frac{L_1(\mathbf{x})}{\|L_1(\mathbf{x})\|}\right)\right\| \leq \|L_2\|$$

since $L_1(\mathbf{x})/\|L_1(\mathbf{x})\| \in \overline{\mathbb{B}}^m$, and thus

$$\|L_2(L_1(\mathbf{x}))\| \leq \|L_2\|\|L_1(\mathbf{x})\| \leq \|L_2\|\|L_1\|.$$

If $L_1(\mathbf{x}) = \mathbf{0}$, then we apply Proposition 10.20(1) to obtain $\|L_2(L_1(\mathbf{x}))\| = 0 \leq \|L_2\|\|L_1\|$ once more. Hence $\|L_2\|\|L_1\|$ is an upper bound for the set

$$\left\{\|L_2(L_1(\mathbf{x}))\| : x \in \overline{\mathbb{B}}^n\right\},$$

and therefore

$$\|L_2 \circ L_1\| = \sup_{\mathbf{x}\in\overline{\mathbb{B}}^n} \|L_2(L_1(\mathbf{x}))\| \leq \|L_2\|\|L_1\|,$$

as desired. ∎

Remark. Another useful fact about $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is that $\|L(\mathbf{x})\| \leq \|L\|\|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$. This is clearly true when $\mathbf{x} = \mathbf{0}$: recalling Proposition 10.20(1) and Theorem 10.33(1), we simply obtain $\mathbf{0} = \mathbf{0}$. If $\mathbf{x} \neq \mathbf{0}$, then

$$\|L(\mathbf{x})\| = \left\|L\left(\|\mathbf{x}\| \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\right\| = \left\|L\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\right\| \|\mathbf{x}\| \leq \|L\|\|\mathbf{x}\|$$

since $\mathbf{x}/\|\mathbf{x}\| \in \overline{\mathbb{B}}^n$.

For each $n \in \mathbb{N}$ the **general linear group of degree n**, denoted by $\mathrm{GL}_n(\mathbb{R})$, is the set of all invertible linear operators on $\mathbb{R}^n$. That is,

$$\mathrm{GL}_n(\mathbb{R}) = \{L \in \mathcal{L}(\mathbb{R}^n) : L \text{ is invertible}\}.$$

In light of Theorem 10.26 and Proposition 10.21, we may write

$$\mathrm{GL}_n(\mathbb{R}) = \left\{L \in \mathcal{L}(\mathbb{R}^n) : \mathrm{Nul}(L) = \{\mathbf{0}\}\right\}.$$

It is left as an exercise to show that $\mathrm{GL}_n(\mathbb{R})$ is a group under the "multiplication" operation of function composition $\circ$ given by Definition 10.22.

Theorem 10.34.

1. *Let* $L \in \mathrm{GL}_n(\mathbb{R})$ *and* $\Lambda \in \mathcal{L}(\mathbb{R}^n)$. *If*

$$\|\Lambda - L\|\|L^{-1}\| < 1,$$

then $\Lambda \in \mathrm{GL}_n(\mathbb{R})$.

2. $\mathrm{GL}_n(\mathbb{R})$ *is an open subset of* $(\mathcal{L}(\mathbb{R}^n), d_{\sup})$.

3. *The mapping* $\text{Inv} : (\text{GL}_n(\mathbb{R}), d_{\sup}) \to (\text{GL}_n(\mathbb{R}), d_{\sup})$ *given by* $\text{Inv}(L) = L^{-1}$ *is a homeomorphism.*

Proof.
Proof of Part (1). Suppose $\Lambda \notin \text{GL}_n(\mathbb{R})$. Then Λ is not a bijection by Theorem 10.26, which is to say Λ is either not injective or not surjective. By Proposition 10.21 it follows that $\text{Nul}(\Lambda) \neq \{\mathbf{0}\}$, and so there exists some $\mathbf{x}_0 \neq \mathbf{0}$ such that $\Lambda(\mathbf{x}_0) = \mathbf{0}$. Let $\hat{\mathbf{x}}_0 = \mathbf{x}_0/\|\mathbf{x}_0\|$, so $\hat{\mathbf{x}}_0 \in \overline{\mathbb{B}}^n$ with $\|\hat{\mathbf{x}}_0\| = 1$. Noting that $\text{Nul}(L) = \{\mathbf{0}\}$, we have $L(\hat{\mathbf{x}}_0) = \mathbf{y}_0 \neq \mathbf{0}$. Now,

$$\|(\Lambda - L)(\hat{\mathbf{x}}_0)\| = \|0 - L(\hat{\mathbf{x}}_0)\| = \|L(\hat{\mathbf{x}}_0)\| = \|\mathbf{y}_0\|$$

shows that $\|\Lambda - L\| \geq \|\mathbf{y}_0\| > 0$. Also

$$\left\| L^{-1}\left(\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right)\right\| = \frac{\|L^{-1}(\mathbf{y}_0)\|}{\|\mathbf{y}_0\|} = \frac{\|\hat{\mathbf{x}}_0\|}{\|\mathbf{y}_0\|} = \frac{1}{\|\mathbf{y}_0\|}$$

shows that $\|L^{-1}\| \geq 1/\|\mathbf{y}_0\|$. Thus

$$\|\Lambda - L\|\|L^{-1}\| \geq \|\mathbf{y}_0\|\left(\frac{1}{\|\mathbf{y}_0\|}\right) = 1.$$

Proof of Part (2). Fix $L_0 \in \text{GL}_n(\mathbb{R})$, so $\|L_0^{-1}\| > 0$. Let $L \in \mathcal{L}(\mathbb{R}^n)$ be such that

$$\|L - L_0\| < \frac{1}{\|L_0^{-1}\|}.$$

Then $\|L - L_0\|\|L_0^{-1}\| < 1$, and so by part (1) it follows that $L \in \text{GL}_n(\mathbb{R})$. Thus the open ball $B_\epsilon(L_0)$ is a subset of $\text{GL}_n(\mathbb{R})$ for $\epsilon = 1/\|L_0^{-1}\|$. This shows that every point in $\text{GL}_n(\mathbb{R})$ is an interior point, and therefore $\text{GL}_n(\mathbb{R})$ is an open set.

Proof of Part (3). It is clear that the mapping Inv is a bijection. Fix $L_0 \in \text{GL}_n(\mathbb{R})$, and let $\epsilon > 0$. Choose

$$\delta = \frac{\epsilon}{\|L^{-1}\|\|L_0^{-1}\|},$$

and suppose that $\|L - L_0\| < \delta$. Now,

$$\|L - L_0\| < \frac{\epsilon}{\|L^{-1}\|\|L_0^{-1}\|} \quad \Rightarrow \quad \|L^{-1}\|\|L - L_0\|\|L_0^{-1}\| < \epsilon,$$

and so by Theorem 10.33(4) we obtain

$$\|L^{-1} - L_0^{-1}\| = \|L^{-1} \circ (L_0 - L) \circ L_0^{-1}\| \leq \|L^{-1}\|\|L_0 - L\|\|L_0^{-1}\| < \epsilon,$$

observing that $\|L_0 - L\| = \|L - L_0\|$ by Theorem 10.33(2). Hence Inv is continuous, and since $\text{Inv}^{-1} = \text{Inv}$ we conclude that Inv is a homeomorphism. ■

Proposition 10.35. *Let* (X, d) *be a metric space, and for each* $1 \leq i \leq m$, $1 \leq j \leq n$ *let* $a_{ij} : X \to \mathbb{R}$ *be a continuous function. Let* $\mathcal{E}_n = (\mathbf{e}_j)_{j=1}^n$ *and* $\mathcal{E}_m = (\boldsymbol{\epsilon}_i)_{i=1}^m$ *denote the standard bases for* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*, respectively. If* $\Theta : (X, d) \to (\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), d_{\sup})$ *is given by* $\Theta(p) = L_p$*, where* $[L_p]_{\mathcal{E}_n\mathcal{E}_m} = [a_{ij}(p)]_{m,n}$*, then* Θ *is continuous on* X*.*

Proof. Fix $\hat{p} \in X$. Let $\epsilon > 0$ be arbitrary. Choose $\delta > 0$ such that $p \in X$ with $d(p, \hat{p}) < \delta$ implies

$$|a_{ij}(p) - a_{ij}(\hat{p})| < \frac{\epsilon}{\sqrt{mn}}$$

for all $1 \leq i \leq m$ and $1 \leq j \leq n$. Fix $p \in X$ with $d(p, \hat{p}) < \delta$, and let $L_p = \Theta(p)$ and $L_{\hat{p}} = \Theta(\hat{p})$, so that $[L_p]_{\mathcal{E}_n\mathcal{E}_m} = [a_{ij}(p)]_{m,n}$ and $[L_{\hat{p}}]_{\mathcal{E}_n\mathcal{E}_m} = [a_{ij}(\hat{p})]_{m,n}$. Recalling (10.7), for each $\mathbf{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$ we have

$$L_p(\mathbf{x}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}(p)x_j\right)\boldsymbol{\epsilon}_i \quad \text{and} \quad L_{\hat{p}}(\mathbf{x}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}(\hat{p})x_j\right)\boldsymbol{\epsilon}_i.$$

Hence, recalling the Schwarz Inequality,

$$\begin{aligned}
\|(L_p - L_{\hat{p}})(\mathbf{x})\|^2 &= \sum_{i=1}^{m}\left(\sum_{j=1}^{n}\big[a_{ij}(p) - a_{ij}(\hat{p})\big]x_j\right)^2 \\
&\leq \sum_{i=1}^{m}\left[\left(\sum_{j=1}^{n}\big[a_{ij}(p) - a_{ij}(\hat{p})\big]^2\right)\left(\sum_{j=1}^{n} x_j^2\right)\right]^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\big[a_{ij}(p) - a_{ij}(\hat{p})\big]^2\|\mathbf{x}\|^2
\end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^n$. Thus

$$\begin{aligned}
d_{\sup}\big(\Theta(p), \Theta(\hat{p})\big) &= \|L_p - L_{\hat{p}}\| = \sup_{\mathbf{x}\in\overline{\mathbb{B}}^n}\|(L_p - L_{\hat{p}})(\mathbf{x})\| \\
&= \sup_{\mathbf{x}\in\overline{\mathbb{B}}^n}\left[\|\mathbf{x}\|\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\big[a_{ij}(p) - a_{ij}(\hat{p})\big]^2\right)^{1/2}\right] \\
&\leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\big[a_{ij}(p) - a_{ij}(\hat{p})\big]^2\right)^{1/2} < \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\epsilon^2}{mn}\right)^{1/2} = \epsilon.
\end{aligned}$$

Therefore Θ is continuous at $\hat{p}$, and we conclude that Θ is continuous on X. ∎

10.4 – The Total Derivative

Definition 10.36. *A mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable at** $\mathbf{x}$ if there is an open set $U \subseteq \mathbb{R}^n$ with $\mathbf{x} \in U \subseteq S$, and linear mapping $\mathrm{d}F_{\mathbf{x}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, such that*

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0, \tag{10.10}$$

*in which case we call $\mathrm{d}F_{\mathbf{x}}$ the **total derivative** (or **differential**) of F at $\mathbf{x}$. We say F is **differentiable on U** if it is differentiable at $\mathbf{x}$ for all $\mathbf{x} \in U$.*

Proposition 10.37. *If $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$, then the linear mapping $\mathrm{d}F_{\mathbf{x}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ for which (10.10) holds is unique.*

Proof. Suppose $L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ are such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_1(\mathbf{h})\|}{\|\mathbf{h}\|} = 0 \quad \text{and} \quad \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

For all $\mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ we have, by the Triangle Inequality,

$$\begin{aligned}\frac{\|L_1(\mathbf{h}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} &\le \frac{\|-F(\mathbf{x}+\mathbf{h}) + F(\mathbf{x}) + L_1(\mathbf{h}) + F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} \\ &\le \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_1(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|},\end{aligned}$$

and since

$$\lim_{\mathbf{h}\to\mathbf{0}} \left(\frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_1(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} \right) = 0,$$

the Squeeze Theorem implies that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|L_1(\mathbf{h}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

Thus for fixed $\mathbf{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ we have

$$\lim_{t\to 0} \frac{\|L_1(t\mathbf{y}) - L_2(t\mathbf{y})\|}{\|t\mathbf{y}\|} = 0,$$

but also we have

$$\begin{aligned}\lim_{t\to 0} \frac{\|L_1(t\mathbf{y}) - L_2(t\mathbf{y})\|}{\|t\mathbf{y}\|} &= \lim_{t\to 0} \frac{|t|\big(\|L_1(\mathbf{y}) - L_2(\mathbf{y})\|\big)}{|t|\|\mathbf{y}\|} = \lim_{t\to 0} \frac{\|L_1(\mathbf{y}) - L_2(\mathbf{y})\|}{\|\mathbf{y}\|} \\ &= \frac{\|L_1(\mathbf{y}) - L_2(\mathbf{y})\|}{\|\mathbf{y}\|}.\end{aligned}$$

It follows that

$$\frac{\|L_1(\mathbf{y}) - L_2(\mathbf{y})\|}{\|\mathbf{y}\|} = 0,$$

and hence $L_1(\mathbf{y}) = L_2(\mathbf{y})$ for all $\mathbf{y} \neq \mathbf{0}$. Since $L_1(\mathbf{0}) = L_2(\mathbf{0}) = \mathbf{0}$ by Proposition 10.20(1), we conclude that $L_1 = L_2$. ■

We introduce a bit of notation for the next proposition. Given a set A in a vector space V, define for any $\mathbf{v} \in V$ the set

$$A + \mathbf{v} = \{\mathbf{a} + \mathbf{v} : \mathbf{a} \in A\}.$$

Also define $A - \mathbf{v} = A + (-\mathbf{v})$.

Proposition 10.38. *A mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$ if and only if there is an open set $U \subseteq \mathbb{R}^n$ with $\mathbf{x} \in U \subseteq S$, some $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and a mapping $R : \mathbb{R}^n \to \mathbb{R}^m$, such that*

$$R(\mathbf{h}) = F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - L(\mathbf{h}) \tag{10.11}$$

for all $\mathbf{h} \in U - \mathbf{x}$ and

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|R(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

Proof. Suppose $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$. Then there is an open set $U \subseteq \mathbb{R}^n$ with $\mathbf{x} \in U \subseteq S$, and a linear mapping $\mathrm{d}F_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^m$, such that (10.10) holds. Define $R : \mathbb{R}^n \to \mathbb{R}^m$ by

$$R(\mathbf{h}) = F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})$$

for all $\mathbf{h} \in U - \mathbf{x}$. Since U is open, there exists some $\epsilon > 0$ such that $\mathbf{x} + \mathbf{h} \in U$ for all $\mathbf{h} \in \mathbb{R}^n$ with $\|\mathbf{h}\| < \epsilon$, which is to say $\mathbf{h} \in U - \mathbf{x}$ and so $B_\epsilon(\mathbf{0}) \subseteq U - \mathbf{x}$. We then obtain

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|R(\mathbf{h})\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \to \mathbf{0}} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0,$$

as desired.

Now suppose there is an open set $U \subseteq \mathbb{R}^n$ with $\mathbf{x} \in U \subseteq S$, some $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and a mapping $R : \mathbb{R}^n \to \mathbb{R}^m$, such that (10.11) holds for all $\mathbf{h} \in U - \mathbf{x}$ and $\|R(\mathbf{h})\|/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$. Then

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - L(\mathbf{h})\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \to \mathbf{0}} \frac{\|R(\mathbf{h})\|}{\|\mathbf{h}\|} = 0,$$

and so F is differentiable at $\mathbf{x}$ by Definition 10.36 (with $L = \mathrm{d}F_{\mathbf{x}}$ by Proposition 10.37). ∎

In Proposition 10.38 it is immaterial how the mapping R is defined outside $U - \mathbf{x}$, and so we may simply set $R \equiv \mathbf{0}$ on $\mathbb{R}^n \setminus (U - \mathbf{x})$. The alternate characterization of differentiability established by the proposition is frequently given as the *definition* of differentiability in the literature.

Proposition 10.39. *If $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $\mathrm{d}L_{\mathbf{x}} = L$ for all $\mathbf{x} \in \mathbb{R}^n$.*

Proof. Suppose $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and fix $\mathbf{x} \in \mathbb{R}^n$. For any $\mathbf{h} \in \mathbb{R}^n$, since $L(\mathbf{x}+\mathbf{h}) = L(\mathbf{x}) + L(\mathbf{h})$, we have

$$L(\mathbf{x} + \mathbf{h}) - L(\mathbf{x}) - L(\mathbf{h}) = L(\mathbf{x}) + L(\mathbf{h}) - L(\mathbf{x}) - L(\mathbf{h}) = \mathbf{0}$$

and so

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|L(\mathbf{x} + \mathbf{h}) - L(\mathbf{x}) - L(\mathbf{h})\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \to \mathbf{0}} (0) = 0$$

obtains easily. This shows that $\mathrm{d}L_{\mathbf{x}} = L$. ∎

Proposition 10.40. *If $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$, then it is continuous at $\mathbf{x}$.*

Proof. Suppose F is differentiable at $\mathbf{x}$. By Proposition 10.38 there is an open $U \subseteq \mathrm{Dom}(F)$ with $\mathbf{x} \in U$ such that

$$F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) = R(\mathbf{h}) + \mathrm{d}F_{\mathbf{x}}(\mathbf{h})$$

for all $\mathbf{h}$ for which $\mathbf{x}+\mathbf{h} \in U$, where $\|R(\mathbf{h})\|/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$. Fix $\epsilon > 0$. By Theorem 10.33(1) the linear mapping $\mathrm{d}F_{\mathbf{x}}$ is continuous at $\mathbf{0}$, and so there exists some $\delta_1 > 0$ such that $\|\mathbf{h}\| < \delta_1$ implies

$$\|\mathrm{d}F_{\mathbf{x}}(\mathbf{h})\| = \|\mathrm{d}F_{\mathbf{x}}(\mathbf{h}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{0})\| < \frac{\epsilon}{2}.$$

Also $\|R(\mathbf{h})\|/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$ implies that $R(\mathbf{h}) \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$, and so there exists some $\delta_2 > 0$ such that $\|\mathbf{h}\| < \delta_2$ implies $\|R(\mathbf{h})\| < \epsilon/2$. (Observe that $\|\mathbf{h}\| < \delta_2$ also implies that $\mathbf{x}+\mathbf{h} \in U$.) Suppose $\boldsymbol{\xi} \in \mathrm{Dom}(F)$ is such that $\|\boldsymbol{\xi} - \mathbf{x}\| < \min\{\delta_1, \delta_2\}$. Then $\boldsymbol{\xi} \in U$ with

$$\|\mathrm{d}F_{\mathbf{x}}(\boldsymbol{\xi} - \mathbf{x})\| < \frac{\epsilon}{2} \quad \text{and} \quad \|R(\boldsymbol{\xi} - \mathbf{x})\| < \frac{\epsilon}{2},$$

and so

$$\begin{aligned}\|F(\boldsymbol{\xi}) - F(\mathbf{x})\| &= \|F(\mathbf{x} + (\boldsymbol{\xi} - \mathbf{x})) - F(\mathbf{x})\| = \|R(\boldsymbol{\xi} - \mathbf{x}) + \mathrm{d}F_{\mathbf{x}}(\boldsymbol{\xi} - \mathbf{x})\| \\ &\leq \|R(\boldsymbol{\xi} - \mathbf{x})\| + \|\mathrm{d}F_{\mathbf{x}}(\boldsymbol{\xi} - \mathbf{x})\| < \epsilon.\end{aligned}$$

Therefore F is continuous at $\mathbf{x}$. ■

The following Chain Rule can help to determine whether a given mapping H is differentiable at some point in the interior of its domain, and if it is, to then find the total derivative of H at that point. However, it is necessary to characterize H as a composition of two mappings G and F whose differentiability and relevant total derivatives are known.

Theorem 10.41 (Chain Rule). *Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open. If $F : U \to \mathbb{R}^m$ is differentiable at $\mathbf{x} \in U$, $F(U) \subseteq V$, and $G : V \to \mathbb{R}^\ell$ is differentiable at $F(\mathbf{x})$, then $G \circ F$ is differentiable at $\mathbf{x}$ with*

$$\mathrm{d}(G \circ F)_{\mathbf{x}} = \mathrm{d}G_{F(\mathbf{x})} \circ \mathrm{d}F_{\mathbf{x}}.$$

Proof. Suppose $F : U \to \mathbb{R}^m$ is differentiable at $\mathbf{x} \in U$, $F(U) \subseteq V$, and $G : V \to \mathbb{R}^\ell$ is differentiable at $F(\mathbf{x})$. Let

$$R(\mathbf{h}) = F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})$$

for all $\mathbf{h} \in \mathbb{R}^n$ such that $\mathbf{x}+\mathbf{h} \in U$, and let

$$T(\mathbf{k}) = G(F(\mathbf{x}) + \mathbf{k}) - G(F(\mathbf{x})) - \mathrm{d}G_{F(\mathbf{x})}(\mathbf{k})$$

for all $\mathbf{k} \in \mathbb{R}^m$ such that $F(\mathbf{x}) + \mathbf{k} \in V$. Also define

$$\delta(\mathbf{h}) = \frac{\|R(\mathbf{h})\|}{\|\mathbf{h}\|} \quad \text{and} \quad \epsilon(\mathbf{k}) = \frac{\|T(\mathbf{k})\|}{\|\mathbf{k}\|}. \tag{10.12}$$

Then $\lim_{\mathbf{h}\to\mathbf{0}} \delta(\mathbf{h}) = 0$ and $\lim_{\mathbf{k}\to\mathbf{0}} \epsilon(\mathbf{k}) = 0$ by Proposition 10.38.

Now, for each $\mathbf{h} \in \mathbb{R}^n$ for which $\mathbf{x}+\mathbf{h} \in U$, set $\mathbf{k} = F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x})$. Then, recalling the remark after the proof of Theorem 10.33,

$$\|\mathbf{k}\| = \|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x})\| = \|\mathrm{d}F_{\mathbf{x}}(\mathbf{h}) + R(\mathbf{h})\|$$

$$\leq \|\mathrm{d}F_{\mathbf{x}}(\mathbf{h})\| + \|R(\mathbf{h})\| = \|\mathrm{d}F_{\mathbf{x}}\|\|\mathbf{h}\| + \delta(\mathbf{h})\|\mathbf{h}\|, \tag{10.13}$$

and

$$\begin{aligned}(G\circ F)(\mathbf{x}+\mathbf{h}) - (G\circ F)(\mathbf{x}) - (\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}})(\mathbf{h}) \\ &= G(F(\mathbf{x}+\mathbf{h})) - G(F(\mathbf{x})) - (\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}})(\mathbf{h}) \\ &= G(F(\mathbf{x})+\mathbf{k}) - G(F(\mathbf{x})) - (\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}})(\mathbf{h}) \\ &= \mathrm{d}G_{F(\mathbf{x})}\big(\mathbf{k} - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big) + T(\mathbf{k}) \\ &= \mathrm{d}G_{F(\mathbf{x})}(R(\mathbf{h})) + T(\mathbf{k}).\end{aligned}$$

By (10.12) and (10.13) we have, for any $\mathbf{h}\neq\mathbf{0}$,

$$\begin{aligned}\frac{\|(G\circ F)(\mathbf{x}+\mathbf{h}) - (G\circ F)(\mathbf{x}) - (\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}})(\mathbf{h})\|}{\|\mathbf{h}\|} &= \frac{\|\mathrm{d}G_{F(\mathbf{x})}(R(\mathbf{h})) + T(\mathbf{k})\|}{\|\mathbf{h}\|} \\ \leq \frac{\|\mathrm{d}G_{F(\mathbf{x})}(R(\mathbf{h}))\| + \|T(\mathbf{k})\|}{\|\mathbf{h}\|} &\leq \frac{\|\mathrm{d}G_{F(\mathbf{x})}\|\|R(\mathbf{h})\| + \|T(\mathbf{k})\|}{\|\mathbf{h}\|} \\ = \frac{\delta(\mathbf{h})\|\mathrm{d}G_{F(\mathbf{x})}\|\|\mathbf{h}\| + \epsilon(\mathbf{k})\|\mathbf{k}\|}{\|\mathbf{h}\|} &\leq \delta(\mathbf{h})\|\mathrm{d}G_{F(\mathbf{x})}\| + \epsilon(\mathbf{k})\big[\|\mathrm{d}F_{\mathbf{x}}\| + \delta(\mathbf{h})\big].\end{aligned}$$

As $\mathbf{h}\to\mathbf{0}$ we have $\delta(\mathbf{h})\to 0$, and thus $\mathbf{k}\to\mathbf{0}$ by (10.13). Since $\epsilon(\mathbf{k})\to 0$ as $\mathbf{k}\to\mathbf{0}$, we find that

$$\lim_{\mathbf{h}\to\mathbf{0}}\Big(\delta(\mathbf{h})\|\mathrm{d}G_{F(\mathbf{x})}\| + \epsilon(\mathbf{k})\big[\|\mathrm{d}F_{\mathbf{x}}\| + \delta(\mathbf{h})\big]\Big) = 0,$$

and therefore

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|(G\circ F)(\mathbf{x}+\mathbf{h}) - (G\circ F)(\mathbf{x}) - (\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}})(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

Since $\mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}}$ is a linear mapping by Proposition 10.23, we conclude that $G\circ F$ is differentiable at $\mathbf{x}$, with $\mathrm{d}(G\circ F)_{\mathbf{x}} = \mathrm{d}G_{F(\mathbf{x})}\circ \mathrm{d}F_{\mathbf{x}}$. ∎

10.5 – Partial Derivatives

Let $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a mapping, and let $\mathcal{E}_n = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $\mathcal{E}_m = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_m)$ be the standard (ordered) bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. For each $\mathbf{x} \in \mathbb{R}^n$ we have $F(\mathbf{x}) \in \mathbb{R}^m$, and thus

$$F(\mathbf{x}) = \sum_{i=1}^{m} F_i(\mathbf{x})\boldsymbol{\epsilon}_i$$

for some real-valued scalars $F_1(\mathbf{x}), \ldots, F_m(\mathbf{x})$. In the notation of $\mathcal{E}_m$-coordinates,

$$F(\mathbf{x}) = \begin{bmatrix} F_1(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{bmatrix},$$

which we may sometimes write as

$$F(\mathbf{x}) = \big(F_1(\mathbf{x}), \ldots, F_m(\mathbf{x})\big), \tag{10.14}$$

where the functions $F_i : S \subseteq \mathbb{R}^n \to \mathbb{R}$ are called the **components** of F. In the notation of $\mathcal{E}_n$-coordinates each $\mathbf{x} \in \mathbb{R}^n$ is given by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

for some $x_1, \ldots, x_n \in \mathbb{R}$, and for each $1 \leq i \leq m$ we define

$$F_i(x_1, \ldots, x_n) = F_i\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = F_i(\mathbf{x}). \tag{10.15}$$

We see that each F_i is a real-valued function of n real-valued independent variables.

Definition 10.42. *Let $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be as above. For $\mathbf{x} \in \mathrm{Int}(S)$ we define, for each $1 \leq i \leq m$ and $1 \leq j \leq n$,*

$$(\partial_j F_i)(\mathbf{x}) = \lim_{t \to 0} \frac{F_i(\mathbf{x} + t\mathbf{e}_j) - F_i(\mathbf{x})}{t},$$

provided the limit exists.

More explicitly we have

$$(\partial_j F_i)(x_1, \ldots, x_n) = \lim_{t \to 0} \frac{F_i(x_1, \ldots, x_j + t, \ldots, x_n) - F_i(x_1, \ldots, x_n)}{t},$$

which shows the limit to be the derivative of F_i with respect to the variable x_j while keeping the other variables fixed at the values $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$. For this reason we call $\partial_j F_i$ the **partial derivative of $\boldsymbol{F_i}$ with respect to $\boldsymbol{x_j}$**, also denoted by $\partial_{x_j} F_i$.

Theorem 10.43. *Let* $\mathcal{E}_n = (\mathbf{e}_j)_{j=1}^n$ *and* $\mathcal{E}_m = (\boldsymbol{\epsilon}_i)_{i=1}^m$ *be the standard bases for* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*. If* $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\mathbf{x}$*, then all partial derivatives* $(\partial_j F_i)(\mathbf{x})$ *exist, and*

$$\mathrm{d}F_{\mathbf{x}}(\mathbf{e}_j) = \sum_{i=1}^m (\partial_j F_i)(\mathbf{x})\boldsymbol{\epsilon}_i$$

for all $1 \leq j \leq n$*.*

Proof. Suppose $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$. Letting $(\mathrm{d}F_{\mathbf{x}})_i$ denote the component functions of $\mathrm{d}F_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^m$, so that

$$\mathrm{d}F_{\mathbf{x}}(\mathbf{h}) = \sum_{i=1}^m (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{h})\boldsymbol{\epsilon}_i$$

for each $\mathbf{h} \in \mathbb{R}^n$, we have

$$\begin{aligned} 0 &= \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} \\ &= \lim_{\mathbf{h}\to\mathbf{0}} \left\| \sum_{i=1}^m \frac{F_i(\mathbf{x}+\mathbf{h}) - F_i(\mathbf{x}) - (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{h})}{\|\mathbf{h}\|}\boldsymbol{\epsilon}_i \right\| \\ &= \lim_{\mathbf{h}\to\mathbf{0}} \left[\sum_{i=1}^m \left(\frac{F_i(\mathbf{x}+\mathbf{h}) - F_i(\mathbf{x}) - (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{h})}{\|\mathbf{h}\|} \right)^2 \right]^{1/2}, \end{aligned}$$

and so

$$\lim_{\mathbf{h}\to\mathbf{0}} \sum_{i=1}^m \left(\frac{F_i(\mathbf{x}+\mathbf{h}) - F_i(\mathbf{x}) - (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{h})}{\|\mathbf{h}\|} \right)^2 = 0.$$

Thus for any $1 \leq j \leq n$ we have

$$\lim_{t\to 0} \sum_{i=1}^m \left(\frac{F_i(\mathbf{x}+t\mathbf{e}_j) - F_i(\mathbf{x}) - (\mathrm{d}F_{\mathbf{x}})_i(t\mathbf{e}_j)}{\|t\mathbf{e}_j\|} \right)^2 = 0,$$

and hence

$$\lim_{t\to 0} \sum_{i=1}^m \left(\frac{F_i(\mathbf{x}+t\mathbf{e}_j) - F_i(\mathbf{x}) - t(\mathrm{d}F_{\mathbf{x}})_i(\mathbf{e}_j)}{t} \right)^2 = 0.$$

It follows that, for any $1 \leq i \leq m$,

$$\lim_{t\to 0} \frac{F_i(\mathbf{x}+t\mathbf{e}_j) - F_i(\mathbf{x}) - t(\mathrm{d}F_{\mathbf{x}})_i(\mathbf{e}_j)}{t} = \lim_{t\to 0} \left(\frac{F_i(\mathbf{x}+t\mathbf{e}_j) - F_i(\mathbf{x})}{t} - (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{e}_j) \right) = 0,$$

which implies that

$$(\partial_j F_i)(\mathbf{x}) = \lim_{t\to 0} \frac{F_i(\mathbf{x}+t\mathbf{e}_j) - F_i(\mathbf{x})}{t} = (\mathrm{d}F_{\mathbf{x}})_i(\mathbf{e}_j)$$

for all $1 \leq i \leq m$ and $1 \leq j \leq n$. This shows that all partial derivatives of F exist, and moreover

$$\mathrm{d}F_{\mathbf{x}}(\mathbf{e}_j) = \sum_{i=1}^m (\partial_j F_i)(\mathbf{x})\boldsymbol{\epsilon}_i$$

for all $1 \leq j \leq n$. ■

Given $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$, for any $\mathbf{h} = \sum_{j=1}^n h_j\mathbf{e}_j \in \mathbb{R}^n$ we have

$$\mathrm{d}F_{\mathbf{x}}(\mathbf{h}) = \sum_{j=1}^n h_j \mathrm{d}F_{\mathbf{x}}(\mathbf{e}_j) = \sum_{j=1}^n \left(h_j \sum_{i=1}^m (\partial_j F_i)(\mathbf{x})\boldsymbol{\epsilon}_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n (\partial_j F_i)(\mathbf{x}) h_j \right) \boldsymbol{\epsilon}_i,$$

by Theorem 10.43, and so the $\mathcal{E}_m$-coordinates of $(\mathrm{d}F_{\mathbf{x}})(\mathbf{h}) \in \mathbb{R}^m$ are seen to be

$$\big[\mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big]_{\mathcal{E}_m} = \begin{bmatrix} \sum_{j=1}^n (\partial_j F_1)(\mathbf{x}) h_j \\ \vdots \\ \sum_{j=1}^n (\partial_j F_m)(\mathbf{x}) h_j \end{bmatrix}.$$

Since

$$\begin{bmatrix} (\partial_1 F_1)(\mathbf{x}) & \cdots & (\partial_n F_1)(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ (\partial_1 F_m)(\mathbf{x}) & \cdots & (\partial_n F_m)(\mathbf{x}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n (\partial_j F_1)(\mathbf{x}) h_j \\ \vdots \\ \sum_{j=1}^n (\partial_j F_m)(\mathbf{x}) h_j \end{bmatrix},$$

we see that the matrix corresponding to $\mathrm{d}F_{\mathbf{x}}$ with respect to the bases $\mathcal{E}_n$ and $\mathcal{E}_m$ (what may be called the $\mathcal{E}_n\mathcal{E}_m$-matrix for $\mathrm{d}F_{\mathbf{x}}$) is

$$[\mathrm{d}F_{\mathbf{x}}]_{\mathcal{E}_n\mathcal{E}_m} = \begin{bmatrix} (\partial_1 F_1)(\mathbf{x}) & \cdots & (\partial_n F_1)(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ (\partial_1 F_m)(\mathbf{x}) & \cdots & (\partial_n F_m)(\mathbf{x}) \end{bmatrix}, \tag{10.16}$$

or more compactly

$$[\mathrm{d}F_{\mathbf{x}}] = \big[(\partial_j F_i)(\mathbf{x})\big]_{mn},$$

which is in agreement with (10.8) in the general treatment. That is, $[\mathrm{d}F_{\mathbf{x}}]_{\mathcal{E}_n\mathcal{E}_m}$ is the unique matrix for which

$$[\mathrm{d}F_{\mathbf{x}}]_{\mathcal{E}_n\mathcal{E}_m}[\mathbf{h}]_{\mathcal{E}_n} = \big[\mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big]_{\mathcal{E}_m}$$

holds for all $\mathbf{h} \in \mathbb{R}^n$. Since it is understood that we are working with the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we may simply write

$$[\mathrm{d}F_{\mathbf{x}}]\mathbf{h} = \mathrm{d}F_{\mathbf{x}}(\mathbf{h}),$$

where $\mathbf{h}$ and $(\mathrm{d}F_{\mathbf{x}})(\mathbf{h})$ are column vectors. Commonly the symbol $\mathrm{d}F_{\mathbf{x}}$ is used to denote the matrix $[\mathrm{d}F_{\mathbf{x}}]$, and so $[\mathrm{d}F_{\mathbf{x}}]\mathbf{h}$ is further simplified to $\mathrm{d}F_{\mathbf{x}}\mathbf{h}$.

In the case when $m = n$, which is to say $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^n$, then the matrix (10.16) becomes a square matrix provided that F is differentiable at $\mathbf{x}$. The determinant

$$J_F(\mathbf{x}) = \det\big([\mathrm{d}F_{\mathbf{x}}]_{\mathcal{E}_n\mathcal{E}_n}\big) = \begin{vmatrix} (\partial_1 F_1)(\mathbf{x}) & \cdots & (\partial_n F_1)(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ (\partial_1 F_n)(\mathbf{x}) & \cdots & (\partial_n F_n)(\mathbf{x}) \end{vmatrix} \tag{10.17}$$

is the **Jacobian of $\boldsymbol{F}$ at $\mathbf{x}$**, which may also be denoted by $\det(\mathrm{d}F_{\mathbf{x}})$. If F is differentiable on an open set $U \subseteq S$, then $J_F : U \to \mathbb{R}$.

Partial differentiation of a function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ follows the same rules given in Chapter 5 for the differentiation of a real-valued function of a single real variable, and so (10.16) offers a straightforward means of determining the total derivative of a mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$.

Example 10.44. In the classical study of the differential geometry of surfaces the notation employed may be quite different.[2] A surface $\Sigma \subseteq \mathbb{R}^3$ is typically characterized as the trace of a parametrization $\mathbf{x} : S \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\mathbf{x}\left(\begin{bmatrix} u \\ v \end{bmatrix}\right) = \begin{bmatrix} x(u,v) \\ y(u,v) \\ z(u,v) \end{bmatrix},$$

where here we assume the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$. Note the bold-facing used in the function's symbol, which is not a notational practice in these notes but is prevalent in the literature. Adopting the notation defined by (10.14) and (10.15), we may write

$$\mathbf{x}(u,v) = \big(x(u,v), y(u,v), z(u,v)\big).$$

The differential of $\mathbf{x}$ at a point $p = (u_0, v_0)$ in S is then the linear map $\mathrm{d}\mathbf{x}_p : \mathbb{R}^2 \to \mathbb{R}^3$ with corresponding matrix

$$\mathrm{d}\mathbf{x}_p = \begin{bmatrix} x_u(p) & x_v(p) \\ y_u(p) & y_v(p) \\ z_u(p) & z_v(p) \end{bmatrix},$$

where the symbol $\mathrm{d}\mathbf{x}_p$ now stands for $[\mathrm{d}\mathbf{x}_p]$, and of course $x_u := \partial_u x$, $x_v := \partial_v x$, and so on. A further bastardization of notation, never again paraded in public in these notes, is to write

$$\mathrm{d}\mathbf{x}_p = \begin{bmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{bmatrix},$$

where the evaluation of the partial derivatives at the point p is understood. ■

Example 10.45. Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$F\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x^3 + yz^2 \\ e^{xz} + \sin y \end{bmatrix},$$

which may also be written $F(x,y,z) = (x^3 + yz, e^{xz} + \sin y)$. The components of F are

$$F_1(x,y,z) = x^3 + yz^2 \quad \text{and} \quad F_2(x,y,z) = e^{xz} + \sin y.$$

If F is differentiable at

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3,$$

then the matrix that represents the total derivative of F at $\mathbf{x}$ is

$$[\mathrm{d}F_\mathbf{x}] = \begin{bmatrix} (\partial_1 F_1)(\mathbf{x}) & (\partial_2 F_1)(\mathbf{x}) & (\partial_3 F_1)(\mathbf{x}) \\ (\partial_1 F_2)(\mathbf{x}) & (\partial_2 F_2)(\mathbf{x}) & (\partial_3 F_2)(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 3x^2 & z^2 & 2yz \\ ze^{xz} & \cos y & xe^{xz} \end{bmatrix}.$$

In particular for $\mathbf{a} = [0,0,2]^\top$ we have

$$[\mathrm{d}F_\mathbf{a}] = \begin{bmatrix} 0 & 4 & 0 \\ 2 & 1 & 0 \end{bmatrix}.$$

[2]See for instance *Differential Geometry of Curves and Surfaces* by Manfredo DoCarmo.

Technically this does not show that F is differentiable at $\mathbf{a}$. In order to do that, our only recourse at this time is to verify that the limit

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a}) - \mathrm{d}F_{\mathbf{a}}(\mathbf{h})\|}{\|\mathbf{h}\|}$$

equals zero, which would be arduous at best. Our theory clearly needs further development. ■

In Chapter 5 we defined a vector-valued function $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ to be differentiable at $t \in I$ if the limit

$$\gamma'(t) = \lim_{h\to 0} \frac{\gamma(t+h) - \gamma(t)}{h} \tag{10.18}$$

exists in $\mathbb{R}^n$. This is equivalent to requiring that there exists some $\gamma'(t) \in \mathbb{R}^n$ such that

$$\lim_{h\to 0} \frac{\gamma(t+h) - \gamma(t) - \gamma'(t)h}{h} = \mathbf{0},$$

or alternatively

$$\lim_{h\to 0} \frac{\|\gamma(t+h) - \gamma(t) - \gamma'(t)h\|}{|h|} = 0. \tag{10.19}$$

(Note that it is not a requirement that t be in the interior of the domain of γ.) On the other hand, by Definition 10.36 we say γ is differentiable at $t \in \mathrm{Int}(I)$ if there exists a linear mapping $\mathrm{d}\gamma_t : \mathbb{R} \to \mathbb{R}^n$ such that

$$\lim_{h\to 0} \frac{\|\gamma(t+h) - \gamma(t) - \mathrm{d}\gamma_t(h)\|}{|h|} = 0. \tag{10.20}$$

Comparing (10.19) and (10.20), and assuming $t \in \mathrm{Int}(I)$, we see that $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ is differentiable at t in the sense of (10.18) if and only if it is differentiable at t in the sense of Definition 10.36, with the associated linear mapping (i.e. total derivative) being the map $\mathrm{d}\gamma_t : \mathbb{R} \to \mathbb{R}^n$ given by $\mathrm{d}\gamma_t(h) = \gamma'(t)h$ for all $h \in \mathbb{R}$. In particular, if it is given that $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ given by

$$\gamma(t) = \sum_{i=1}^{n} \gamma_i(t)\mathbf{e}_i$$

is differentiable in the sense of Definition 10.36 at $t \in \mathrm{Int}(I)$, then Theorem 10.43 may be used to calculate

$$\mathrm{d}\gamma_t(1) = \sum_{i=1}^{n} (\partial_1\gamma_i)(t)\mathbf{e}_i = \sum_{i=1}^{n} \gamma_i'(t)\mathbf{e}_i = \gamma'(t),$$

bearing in mind that $\{1\}$ is the standard basis for $\mathbb{R}$. Hence

$$\mathrm{d}\gamma_t(h) = h\mathrm{d}\gamma_t(1) = h\gamma'(t)$$

for any $h \in \mathbb{R}$. We summarize our findings with a proposition.

Proposition 10.46. *Let $\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^n$ be a vector-valued function, and let $t \in \mathrm{Int}(I)$. Then γ is differentiable in the sense of Definition 10.36 if and only if $\gamma'(t)$ exists in $\mathbb{R}^n$, in which case $\mathrm{d}\gamma_t : \mathbb{R} \to \mathbb{R}^n$ is given by $\mathrm{d}\gamma_t(h) = h\gamma'(t)$ for all $h \in \mathbb{R}$.*

In Proposition 10.46 and the foregoing remarks it is important to distinguish between the *mapping* $\mathrm{d}\gamma_t \in \mathcal{L}(\mathbb{R}, \mathbb{R}^n)$ and the *vector* $\gamma'(t) \in \mathbb{R}^n$.[3]

Theorem 10.47. *Suppose $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on a convex open set $V \subseteq S$, and there exists some $\alpha \in \mathbb{R}$ such that $\|\mathrm{d}F_{\mathbf{x}}\| \le \alpha$ for all $\mathbf{x} \in V$. Then*

$$\|F(\mathbf{x}) - F(\mathbf{y})\| \le \alpha \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}, \mathbf{y} \in V$.

Proof. Fix $\mathbf{x}, \mathbf{y} \in V$, and define $\gamma : \mathbb{R} \to \mathbb{R}^n$ by

$$\gamma(t) = t\mathbf{x} + (1-t)\mathbf{y}.$$

Observe that $\gamma([0,1])$ is the segment $[\mathbf{x}, \mathbf{y}]$ in $\mathbb{R}^n$, and so $\gamma([0,1]) \subseteq V$ since V is convex. Indeed, since γ is continuous and V is open, there exists an open interval $I \subseteq \mathbb{R}$ such that $[0,1] \subseteq I$ and $\gamma(I) \subseteq V$. Since $\gamma : I \to \mathbb{R}^n$ is differentiable at $t \in I$, $\gamma(I) \subseteq V$, and $F : V \to \mathbb{R}^m$ is differentiable at $\gamma(t)$, by the Chain Rule $g = F \circ \gamma$ is differentiable at t with

$$\mathrm{d}g_t = \mathrm{d}F_{\gamma(t)} \circ \mathrm{d}\gamma_t.$$

In particular we see that g is differentiable on I. Fix $t \in [0,1]$. Observing that $\overline{\mathbb{B}}^1 = [-1,1]$ and $\gamma'(t) = \mathbf{x} - \mathbf{y}$, by Theorem 10.33(4) and Proposition 10.46,

$$\|\mathrm{d}g_t\| \le \|\mathrm{d}F_{\gamma(t)}\| \|\mathrm{d}\gamma_t\| \le \alpha \sup_{h \in [-1,1]} \|\mathrm{d}\gamma_t(h)\| = \alpha \sup_{h \in [-1,1]} \|h\gamma'(t)\| = \alpha \|\gamma'(t)\| = \alpha \|\mathbf{x} - \mathbf{y}\|.$$

On the other hand, another application of Proposition 10.46 gives

$$\|\mathrm{d}g_t\| = \sup_{h \in [-1,1]} \|\mathrm{d}g_t(h)\| = \sup_{h \in [-1,1]} \|hg'(t)\| = \|g'(t)\|,$$

and hence

$$\|g'(t)\| \le \alpha \|\mathbf{x} - \mathbf{y}\|$$

for all $t \in [0,1]$.

Now, g is differentiable on $(0,1) \subseteq I$, and by Proposition 10.40 g is continuous on $[0,1] \subseteq I$. Thus by Theorem 5.19 in [Rud] there is some $\tau \in (0,1)$ such that $\|g(1) - g(0)\| \le \|g'(\tau)\|$. Therefore

$$\|F(\mathbf{x}) - F(\mathbf{y})\| = \|F(\gamma(1)) - F(\gamma(0))\| = \|g(1) - g(0)\| \le \|g'(\tau)\| \le \alpha \|\mathbf{x} - \mathbf{y}\|,$$

as desired. ■

Proposition 10.48. *Suppose $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on a connected open set $U \subseteq S$. If $\mathrm{d}F_{\mathbf{x}} = \mathbf{0}$ for all $\mathbf{x} \in U$, then F is constant on U.*

---

[3] Many authors use the same symbol to represent both in introductory texts, which is unfortunate.

Proof. Suppose $\mathrm{d}F_{\mathbf{x}} = \mathbf{0}$ for all $\mathbf{x} \in U$. Assume U is convex. Since $\|\mathrm{d}F_{\mathbf{x}}\| = 0$ for all $\mathbf{x} \in U$, Theorem 10.47 implies that $\|F(\mathbf{x}) - F(\mathbf{y})\| = 0$ for all $\mathbf{x}, \mathbf{y} \in U$, and thus $F(\mathbf{x}) = F(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in U$. Therefore F is constant on U.

Now assume U is not convex. Since U is open, for each $\mathbf{x} \in U$ we may choose some $\epsilon_{\mathbf{x}} > 0$ such that $B_{\epsilon_{\mathbf{x}}}(\mathbf{x}) \subseteq U$, and thereby construct a collection of open balls

$$\mathcal{B} = \left\{ B_{\epsilon_{\mathbf{x}}}(\mathbf{x}) : \mathbf{x} \in U \right\}$$

such that $\bigcup \mathcal{B} = U$. Open balls are convex sets, and so F is constant on each $B \in \mathcal{B}$. Fix $\mathbf{a} \in U$, let V_1 be the union of all $B \in \mathcal{B}$ on which $F \equiv F(\mathbf{a})$:

$$V_1 = \bigcup \left\{ B \in \mathcal{B} : F \equiv F(\mathbf{a}) \text{ on } B \right\}.$$

Also let

$$V_2 = \bigcup \left\{ B \in \mathcal{B} : F \equiv \mathbf{c} \text{ on } B \text{ for some } \mathbf{c} \neq F(\mathbf{a}) \right\}.$$

Clearly $V_1 \cup V_2 = U$ and $V_1 \cap V_2 = \varnothing$. Also $V_1 \neq \varnothing$ since $B_{\epsilon_{\mathbf{a}}}(\mathbf{a}) \subseteq V_1$. Since U is connected, it follows that $V_2 = \varnothing$. Therefore $U = V_1$, which shows that every $\mathbf{x} \in U$ lies in a ball on which $F \equiv F(\mathbf{a})$, and therefore $F \equiv F(\mathbf{a})$ on U. ■

A mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on some open set $U \subseteq S$ if and only if the total derivative $\mathrm{d}F_{\mathbf{x}}$ is defined for each $\mathbf{x} \in U$. That is, for each $\mathbf{x} \in U$ there exists some $\mathrm{d}F_{\mathbf{x}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and so there is a mapping $\mathbf{x} \mapsto \mathrm{d}F_{\mathbf{x}}$ defined on U. We denote this mapping by $\mathrm{d}F$, so in explicit terms $\mathrm{d}F : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is given by $\mathrm{d}F(\mathbf{x}) = \mathrm{d}F_{\mathbf{x}}$.

Definition 10.49. *A mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is* ***continuously differentiable*** *on an open set $U \subseteq S$ if F is differentiable on U and the mapping $\mathrm{d}F : (U, d_{\mathrm{e}}) \to (\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), d_{\sup})$ is continuous on U.*

The collection of all mappings that are continuously differentiable on U is denoted by $\mathcal{C}'(U)$, and any $F \in \mathcal{C}'(U)$ is called a ***$\mathcal{C}'$-mapping on U***.

The following result extends the Mean Value Theorem to a multivariable setting, and will be useful in proving the theorem that comes after.

Theorem 10.50 (Multivariable Mean Value Theorem). *Suppose $f : B_r(\mathbf{a}) \subseteq \mathbb{R}^n \to \mathbb{R}$ has continuous first partials. Then for any $\mathbf{x} \in B_r(\mathbf{a})$ there exist $\mathbf{c}_1, \ldots, \mathbf{c}_n \in B_r(\mathbf{a})$ such that*

$$f(\mathbf{x}) - f(\mathbf{a}) = \sum_{i=1}^{n} \partial_i f(\mathbf{c}_i)(\mathbf{x} - \mathbf{a}) \cdot \mathbf{e}_i.$$

Proof. Fix $\mathbf{x} \in B_r(\mathbf{a})$. Thus we have $\mathbf{a} = \sum_{i=1}^n a_i \mathbf{e}_i$ and $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$. For each $1 \leq i \leq n$ let

$$\mathbf{b}_i = \mathbf{a} + \sum_{k=1}^{i} (x_k - a_k)\mathbf{e}_k.$$

Defining $\mathbf{b}_0 = \mathbf{a}$, set $S_i = [\mathbf{b}_{i-1}, \mathbf{b}_i]$ and

$$\Delta_i = \left[\min\{a_i, x_i\}, \max\{a_i, x_i\} \right]$$

for each $1 \leq i \leq n$. It is straightforward to verify that $\mathbf{b}_n = \mathbf{x}$ and $S = \bigcup_{i=1}^n S_i \subseteq B_r(\mathbf{a})$. Fix $1 \leq i \leq n$, and assume that $a_i \neq x_i$. Define $g_i : \Delta_i \to \mathbb{R}$ by

$$g_i(t) = f\big(\mathbf{b}_{i-1} + (t - a_i)\mathbf{e}_i\big) = f(x_1, \ldots, x_{i-1}, t, a_{i+1}, \ldots, a_n).$$

Clearly g_i is continuous on Δ_i. For any $t \in \operatorname{Int}(\Delta_i)$, letting $\mathbf{v}_{it} = \mathbf{b}_{i-1} + (t - a_i)\mathbf{e}_i$, we have

$$g_i'(t) = \lim_{h\to 0} \frac{g_i(t+h) - g_i(t)}{h} = \lim_{h\to 0} \frac{f(\mathbf{v}_{it} + h\mathbf{e}_i) - f(\mathbf{v}_{it})}{h} = \partial_i f(\mathbf{v}_{it}),$$

and so we see that $g_i'(t)$ exists since $\mathbf{v}_{it} \in B_r(\mathbf{a})$ and the partial derivatives of f exist on $B_r(\mathbf{a})$. Hence g_i is differentiable on $\operatorname{Int}(\Delta_i)$. By the Mean Value Theorem it follows that there exists some $t_i \in \operatorname{Int}(\Delta_i)$ such that

$$g_i'(t_i) = \frac{g_i(x_i) - g_i(a_i)}{x_i - a_i}.$$

Setting $\mathbf{c}_i = \mathbf{v}_{it_i}$, we conclude that

$$\partial_i f(\mathbf{c}_i) = \frac{f(\mathbf{b}_i) - f(\mathbf{b}_{i-1})}{x_i - a_i}$$

for all $1 \leq i \leq n$ for which $a_i \neq x_i$. If $a_i = x_i$ simply choose $\mathbf{c}_i = \mathbf{a}_i$. Now,

$$f(\mathbf{x}) - f(\mathbf{a}) = \sum_{i=1}^n \big[f(\mathbf{b}_i) - f(\mathbf{b}_{i-1})\big] = \sum_{i=1}^n \partial_i f(\mathbf{c}_i)(x_i - a_i),$$

observing that for any i for which $a_i = x_i$ we have $\mathbf{b}_{i-1} = \mathbf{b}_i$, and hence

$$f(\mathbf{b}_i) - f(\mathbf{b}_{i-1}) = 0 = \partial_i f(\mathbf{a}_i)(a_i - a_i) = \partial_i f(\mathbf{c}_i)(x_i - a_i).$$

This finishes the proof. ∎

The next theorem will at last provide a practical means of determining whether a mapping is differentiable at a given point in the interior of its domain. For the proof, recall that the standard basis $(\boldsymbol{\epsilon}_i)_{i=1}^m$ for $\mathbb{R}^m$ is orthonormal, so that $\boldsymbol{\epsilon}_i \cdot \boldsymbol{\epsilon}_j := \boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}_j = \delta_{ij}$.

Theorem 10.51. *Let $U \subseteq \mathbb{R}^n$ be an open set, and let $F : U \to \mathbb{R}^m$. Then $F \in \mathcal{C}'(U)$ if and only if the partial derivatives $\partial_j F_i$ exist and are continuous on U for all $1 \leq i \leq m$ and $1 \leq j \leq n$.*

Proof. Suppose $F \in \mathcal{C}'(U)$. Fix $\mathbf{x} \in U$, and let $\epsilon > 0$. Since $\mathrm{d}F$ is continuous at $\mathbf{x}$, there exists some $\delta > 0$ such that $\|\mathrm{d}F_{\boldsymbol{\xi}} - \mathrm{d}F_{\mathbf{x}}\| < \epsilon$ for all $\boldsymbol{\xi} \in U$ with $\|\boldsymbol{\xi} - \mathbf{x}\| < \delta$. Fix $1 \leq i \leq m$ and $1 \leq j \leq n$. By Theorem 10.43,

$$\mathrm{d}F_{\mathbf{x}}(\mathbf{e}_j) \cdot \boldsymbol{\epsilon}_i = \sum_{i=k}^m \partial_j F_k(\mathbf{x}) \boldsymbol{\epsilon}_k \cdot \boldsymbol{\epsilon}_i = \partial_j F_i(\mathbf{x}).$$

Let $\boldsymbol{\xi} \in U$ with $\|\boldsymbol{\xi} - \mathbf{x}\| < \delta$. Then

$$\partial_j F_i(\boldsymbol{\xi}) - \partial_j F_i(\mathbf{x}) = \big[\mathrm{d}F_{\boldsymbol{\xi}}(\mathbf{e}_j) - \mathrm{d}F_{\mathbf{x}}(\mathbf{e}_j)\big] \cdot \boldsymbol{\epsilon}_i = \big[(\mathrm{d}F_{\boldsymbol{\xi}} - \mathrm{d}F_{\mathbf{x}})(\mathbf{e}_j)\big] \cdot \boldsymbol{\epsilon}_i,$$

and since $\|\boldsymbol{\epsilon}_i\| = \|\mathbf{e}_j\| = 1$ (in particular $\mathbf{e}_j \in \overline{\mathbb{B}}^n$), we have

$$\big|\partial_j F_i(\boldsymbol{\xi}) - \partial_j F_i(\mathbf{x})\big| \leq \|(\mathrm{d}F_{\boldsymbol{\xi}} - \mathrm{d}F_{\mathbf{x}})(\mathbf{e}_j)\| \|\boldsymbol{\epsilon}_i\| \leq \|\mathrm{d}F_{\boldsymbol{\xi}} - \mathrm{d}F_{\mathbf{x}}\| < \epsilon.$$

Therefore $\partial_j F_i$ is continuous at $\mathbf{x}$.

For the converse, suppose the partial derivatives $\partial_j F_i$ exist and are continuous on U for all $1 \le i \le m$ and $1 \le j \le n$. Fix $\mathbf{x} \in U$ and let $\epsilon > 0$. Let $r > 0$ be such that $B_r(\mathbf{x}) \subseteq U$. For each $1 \le i \le m$ and $1 \le j \le n$ there exists some $0 < \delta_{ij} < r$ such that $\|\boldsymbol{\xi} - \mathbf{x}\| < \delta_{ij}$ implies

$$\|\partial_j F_i(\boldsymbol{\xi}) - \partial_j F_i(\mathbf{x})\| < \frac{\epsilon}{2mn}.$$

Choose $\delta = \min\{\delta_{ij} : 1 \le i \le m, 1 \le j \le n\}$, and suppose $\|\mathbf{h}\| < \delta$. Fix $1 \le i \le m$. Since $\mathbf{x} + \mathbf{h} \in B_\delta(\mathbf{x}) \subseteq U$, by Theorem 10.50 there exist $\mathbf{c}_{i1}, \dots, \mathbf{c}_{in} \in B_\delta(\mathbf{x})$ such that

$$F_i(\mathbf{x} + \mathbf{h}) - F_i(\mathbf{x}) = \sum_{j=1}^n \partial_j F_i(\mathbf{c}_{ij})\mathbf{h} \cdot \mathbf{e}_j.$$

Now, noting that $|\mathbf{h} \cdot \mathbf{e}_j| \le \|\mathbf{h}\|\|\mathbf{e}_j\| = \|\mathbf{h}\|$, and also $\|\mathbf{c}_{ij} - \mathbf{x}\| < \delta_{ij}$ for each $1 \le j \le n$,

$$\begin{aligned}
\left|\frac{F_i(\mathbf{x} + \mathbf{h}) - F_i(\mathbf{x}) - \sum_{j=1}^n \partial_j F_i(\mathbf{x})\mathbf{h} \cdot \mathbf{e}_j}{\|\mathbf{h}\|}\right| &= \frac{\left|\sum_{j=1}^n \big(\partial_j F_i(\mathbf{c}_{ij}) - \partial_j F_i(\mathbf{x})\big)\mathbf{h} \cdot \mathbf{e}_j\right|}{\|\mathbf{h}\|} \\
&\le \frac{1}{\|\mathbf{h}\|}\sum_{j=1}^n \big|\partial_j F_i(\mathbf{c}_{ij}) - \partial_j F_i(\mathbf{x})\big||\mathbf{h} \cdot \mathbf{e}_j| \le \sum_{j=1}^n \big|\partial_j F_i(\mathbf{c}_{ij}) - \partial_j F_i(\mathbf{x})\big| \\
&\le \sum_{j=1}^n \frac{\epsilon}{2mn} = \frac{\epsilon}{2m}.
\end{aligned}$$

Defining $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ by

$$L(\mathbf{e}_j) = \sum_{i=1}^m \partial_j F_i(\mathbf{x})\boldsymbol{\epsilon}_i$$

for each $1 \le j \le n$, so that

$$L_i(\mathbf{h}) = \sum_{j=1}^n \partial_j F_i(\mathbf{x})\mathbf{h} \cdot \mathbf{e}_j$$

for each $1 \le i \le m$, we have

$$\begin{aligned}
\frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - L(\mathbf{h})\|}{\|\mathbf{h}\|} &= \left[\sum_{i=1}^m \left(\frac{F_i(\mathbf{x} + \mathbf{h}) - F_i(\mathbf{x}) - L_i(\mathbf{h})}{\|\mathbf{h}\|}\right)^2\right]^{1/2} \\
&\le \left(\sum_{i=1}^m \frac{\epsilon^2}{4m^2}\right)^{1/2} = \frac{\epsilon}{2\sqrt{m}} < \epsilon.
\end{aligned}$$

Therefore

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - L(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$$

for the linear mapping L, and we conclude that F is differentiable at $\mathbf{x}$. Since $\mathbf{x} \in U$ is arbitrary, it follows that F is differentiable on U.

Finally, since $\partial_j F_i$ is a real continuous functions on $U \subseteq \mathbb{R}^n$ for each $1 \le i \le m$ and $1 \le j \le n$, and by (10.16) the $\mathcal{E}_n\mathcal{E}_m$-matrix of $\mathrm{d}F_{\mathbf{x}}$ is $[(\partial_j F_i)(\mathbf{x})]$, we conclude by Proposition 10.35 that the mapping $\mathrm{d}F : (U, d_{\mathrm{e}}) \to (\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), d_{\sup})$ given by $\mathrm{d}F(\mathbf{x}) = \mathrm{d}F_{\mathbf{x}}$ is continuous on U. Therefore $F \in \mathcal{C}'(U)$. ■

10.6 – The Inverse Function Theorem

Definition 10.52. *Let (X, d) be a metric space. A mapping $\varphi : X \to X$ is a* ***contraction*** *of X into X if there exists some $c < 1$ such that*

$$d\big(\varphi(x), \varphi(y)\big) \leq cd(x, y)$$

for all $x, y \in X$.

Proposition 10.53 (Contraction Principle). *If (X, d) is a complete metric space and $\varphi : X \to X$ is a contraction, then there exists a unique $x \in X$ such that $\varphi(x) = x$.*

For the following definition, recall that a mapping $F : U \to V$ is invertible if and only if it is bijective, so that there exists a tranformation $F^{-1} : V \to U$ such that $F \circ F^{-1} = I_V$ and $F^{-1} \circ F = I_U$.

Definition 10.54. *Let $U, V \subseteq \mathbb{R}^n$ be open sets. Then $F : U \to V$ is a* ***$\mathcal{C}'$-diffeomorphism*** *if F is bijective, $F \in \mathcal{C}'(U)$, and $F^{-1} \in \mathcal{C}'(V)$.*

We henceforth adopt the following notation: the symbol $\mathrm{d}F_{\mathbf{x}}^{-1}$ indicates the inverse of the total derivative of F at $\mathbf{x}$; that is, $\mathrm{d}F_{\mathbf{x}}^{-1} := (\mathrm{d}F_{\mathbf{x}})^{-1}$. In contrast the symbol $\mathrm{d}(F^{-1})_{\mathbf{y}}$ denotes the total derivative of F^{-1} at $\mathbf{y}$.

Theorem 10.55 (Inverse Function Theorem). *Let $U \subseteq \mathbb{R}^n$ be open, and suppose the mapping $F : U \to \mathbb{R}^n$ is such that $F \in \mathcal{C}'(U)$, $\mathrm{d}F_{\mathbf{a}}$ is invertible for some $\mathbf{a} \in U$, and $\mathbf{b} = F(\mathbf{a})$. Then there exist open sets $A \subseteq U$, $B \subseteq F(U)$ such that $\mathbf{a} \in A$, $\mathbf{b} \in B$, and $F : A \to B$ is a $\mathcal{C}'$-diffeomorphism. Moreover, for each $\mathbf{y} = F(\mathbf{x}) \in B$,*

$$\mathrm{d}(F^{-1})_{\mathbf{y}} = \mathrm{d}F_{\mathbf{x}}^{-1}. \tag{10.21}$$

Proof. Since $\mathrm{d}F : U \to \mathcal{L}(\mathbb{R}^n)$ is continuous at $\mathbf{a}$, there exists some $r > 0$ such that $A_0 = B_r(\mathbf{a}) \subseteq U$ and

$$\|\mathrm{d}F_{\mathbf{x}} - \mathrm{d}F_{\mathbf{a}}\| < \frac{1}{2\|\mathrm{d}F_{\mathbf{a}}^{-1}\|} := \lambda \tag{10.22}$$

for all $\mathbf{x} \in A_0$.

For each $\mathbf{y} \in \mathbb{R}^n$ define $\Phi_{\mathbf{y}} : U \to \mathbb{R}^n$ by

$$\Phi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} + \mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y} - F(\mathbf{x}))$$

for all $\mathbf{x} \in U$. Since $\mathrm{d}F_{\mathbf{a}}^{-1}$ is a linear mapping by Proposition 10.28, we have

$$\Phi_{\mathbf{y}}(\mathbf{x}) = (I - \mathrm{d}F_{\mathbf{a}}^{-1} \circ F)(\mathbf{x}) + \mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y}),$$

where I is the identity tranformation on $\mathbb{R}^n$. For any $\mathbf{x} \in U$ the total derivative of $\Phi_{\mathbf{y}}$ exists, with

$$\mathrm{d}(\Phi_{\mathbf{y}})_{\mathbf{x}} = I - \mathrm{d}F_{\mathbf{a}}^{-1} \circ \mathrm{d}F_{\mathbf{x}} = \mathrm{d}F_{\mathbf{a}}^{-1} \circ (\mathrm{d}F_{\mathbf{a}} - \mathrm{d}F_{\mathbf{x}})$$

by the Chain Rule, Proposition 10.39, and Proposition 10.30(2). Now, by Theorem 10.33(4) and (10.22),

$$\|\mathrm{d}(\Phi_{\mathbf{y}})_{\mathbf{x}}\| \le \|\mathrm{d}F_{\mathbf{a}}^{-1}\|\|\mathrm{d}F_{\mathbf{a}} - \mathrm{d}F_{\mathbf{x}}\| < \frac{1}{2}$$

for all $\mathbf{x} \in A_0$, and so

$$\|\Phi_{\mathbf{y}}(\mathbf{x}_1) - \Phi_{\mathbf{y}}(\mathbf{x}_2)\| \le \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| \tag{10.23}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in A_0$ by Theorem 10.47. This shows that $\Phi_{\mathbf{y}}$ is a contraction of A_0 into A_0, and if we define $A = B_{r/2}(\mathbf{a}) \subseteq U$, so that $\overline{A} \subseteq A_0$, then $\Phi_{\mathbf{y}}$ is a contraction of the *complete* metric space $\overline{A}$ into $\overline{A}$. By the Contraction Principle $\Phi_{\mathbf{y}}$ has a unique fixed point in $\overline{A}$, and so has at most one fixed point in A. Thus there is at most one $\mathbf{x} \in A$ for which

$$\mathbf{x} + \mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y} - F(\mathbf{x})) = \Phi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x},$$

which yields

$$\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y} - F(\mathbf{x})) = \mathbf{0},$$

and hence $F(\mathbf{x}) = \mathbf{y}$ by Proposition 10.21, since $\mathrm{d}F_{\mathbf{a}}^{-1} \in \mathcal{L}(\mathbb{R}^n)$ is invertible and hence a bijection. Since $\mathbf{y} \in \mathbb{R}^n$ is arbitrary, it follows that F is injective on A.

Let $B = F(A) \subseteq F(U)$, so that $F : A \to B$ is bijective, and of course $\mathbf{b} \in B$. We must show that B is open. Fix $\mathbf{y}_0 \in B$. Then $\mathbf{y}_0 = F(\mathbf{x}_0)$ for some $\mathbf{x}_0 \in A$. Let $\rho > 0$ be such that $K := \overline{B}_\rho(\mathbf{x}_0) \subseteq A$. Fix $\mathbf{y} \in B_{\lambda\rho}(\mathbf{y}_0)$. Then

$$\|\Phi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\| = \|\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y} - F(\mathbf{x}_0))\| = \|\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{y} - \mathbf{y}_0)\| < \|\mathrm{d}F_{\mathbf{a}}^{-1}\|\lambda\rho = \frac{\rho}{2},$$

so by (10.23) we have, for any $\mathbf{x} \in K$,

$$\|\Phi_{\mathbf{y}}(\mathbf{x}) - \mathbf{x}_0\| \le \|\Phi_{\mathbf{y}}(\mathbf{x}) - \Phi_{\mathbf{y}}(\mathbf{x}_0)\| + \|\Phi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\| < \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\| + \frac{\rho}{2} \le \rho,$$

and hence $\Phi_{\mathbf{y}}(\mathbf{x}) \in \mathrm{Int}(K)$. Since (10.23) holds on $A_0 \supseteq A$, it holds on K; then, having shown $\Phi_{\mathbf{y}}(K) \subseteq \mathrm{Int}(K)$, it follows that $\Phi_{\mathbf{y}}$ is a contraction of K into K. By the Contraction Principle there exists some $\mathbf{x} \in K$ such that $\Phi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$, and so $F(\mathbf{x}) = \mathbf{y}$. Hence $\mathbf{y} \in F(K) \subseteq F(A) = B$, and since $\mathbf{y} \in B_{\lambda\rho}(\mathbf{y}_0)$ is arbitrary, we obtain $B_{\lambda\rho}(\mathbf{y}_0) \subseteq B$. Therefore B is open.

Next we show that F^{-1} is differentiable on B and verify (10.21). Fix $\mathbf{y} \in B$. Since $F : A \to B$ is bijective, for any $\mathbf{k}$ such that $\mathbf{y} + \mathbf{k} \in B$ there exist $\mathbf{x}, \mathbf{x} + \mathbf{h} \in A$ (with $\mathbf{h} \ne \mathbf{0}$ being unique) such that $F(\mathbf{x}) = \mathbf{y}$ and $F(\mathbf{x} + \mathbf{h}) = \mathbf{y} + \mathbf{k}$. Moreover (10.22) implies that

$$\|\mathrm{d}F_{\mathbf{x}} - \mathrm{d}F_{\mathbf{a}}\|\|\mathrm{d}F_{\mathbf{a}}^{-1}\| < 1,$$

and so $\mathrm{d}F_{\mathbf{x}}$ is an invertible linear operator on $\mathbb{R}^n$ by Theorem 10.34(1). Now,

$$\begin{aligned} F^{-1}(\mathbf{y} + \mathbf{k}) - F^{-1}(\mathbf{y}) - \mathrm{d}F_{\mathbf{x}}^{-1}(\mathbf{k}) &= \mathbf{h} - \mathrm{d}F_{\mathbf{x}}^{-1}(\mathbf{k}) \\ &= \mathrm{d}F_{\mathbf{x}}^{-1}\big(\mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big) - \mathrm{d}F_{\mathbf{x}}^{-1}\big(F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})\big) \\ &= -\mathrm{d}F_{\mathbf{x}}^{-1}\big(F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big); \end{aligned} \tag{10.24}$$

and by (10.23),

$$\|\mathbf{h}\| - \|\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{k})\| \le \|\mathbf{h} - \mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{k})\| = \big\|\mathbf{h} + \mathrm{d}F_{\mathbf{a}}^{-1}\big(F(\mathbf{x}) - F(\mathbf{x} + \mathbf{h})\big)\big\|$$

$$= \|\Phi_{\mathbf{y}}(\mathbf{x}+\mathbf{h}) - \Phi_{\mathbf{y}}(\mathbf{x})\| \le \frac{\|\mathbf{h}\|}{2},$$

whence $\|\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{k})\| \ge \|\mathbf{h}\|/2$ obtains, yielding

$$\|\mathbf{h}\| \le 2\|\mathrm{d}F_{\mathbf{a}}^{-1}(\mathbf{k})\| \le 2\|\mathrm{d}F_{\mathbf{a}}^{-1}\|\|\mathbf{k}\| = \frac{\|\mathbf{k}\|}{\lambda} \tag{10.25}$$

by the remark following Theorem 10.33.

Let $\epsilon > 0$. Since

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$$

by the differentiability of F at $\mathbf{x}$, there exists some $\delta_0 > 0$ such that $0 < \|\mathbf{h}\| < \delta_0$ implies

$$\frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} < \frac{\epsilon\lambda}{\|\mathrm{d}F_{\mathbf{x}}^{-1}\|}.$$

For $r_0 > 0$ such that $B_{r_0}(\mathbf{y}) \subseteq B$, choose $\delta = \min\{r_0, \lambda\delta_0\}$, and suppose $0 < \|\mathbf{k}\| < \delta$. Then by (10.25) the unique $\mathbf{h} \neq \mathbf{0}$ for which $F(\mathbf{x}+\mathbf{h}) = \mathbf{y}+\mathbf{k}$ is such that

$$0 < \|\mathbf{h}\| \le \frac{\|\mathbf{k}\|}{\lambda} < \frac{\delta}{\lambda} \le \delta_0.$$

Now, recalling (10.24),

$$\begin{aligned}
\frac{\|F^{-1}(\mathbf{y}+\mathbf{k}) - F^{-1}(\mathbf{y}) - \mathrm{d}F_{\mathbf{x}}^{-1}(\mathbf{k})\|}{\|\mathbf{k}\|} &\le \frac{\|\mathrm{d}F_{\mathbf{x}}^{-1}\big(F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\big)}{\lambda\|\mathbf{h}\|} \\
&\le \frac{\|\mathrm{d}F_{\mathbf{x}}^{-1}\|}{\lambda} \cdot \frac{\|F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) - \mathrm{d}F_{\mathbf{x}}(\mathbf{h})\|}{\|\mathbf{h}\|} \\
&< \frac{\|\mathrm{d}F_{\mathbf{x}}^{-1}\|}{\lambda} \cdot \frac{\epsilon\lambda}{\|\mathrm{d}F_{\mathbf{x}}^{-1}\|} = \epsilon.
\end{aligned}$$

Thus

$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{\|F^{-1}(\mathbf{y}+\mathbf{k}) - F^{-1}(\mathbf{y}) - \mathrm{d}F_{\mathbf{x}}^{-1}(\mathbf{k})\|}{\|\mathbf{k}\|} = 0,$$

and so F^{-1} is differentiable at $\mathbf{y}$ with $\mathrm{d}(F^{-1})_{\mathbf{y}} = \mathrm{d}F_{\mathbf{x}}^{-1}$. Since $\mathbf{y} \in B$ is arbitrary, F^{-1} is differentiable on B, and (10.21) is verified.

It is clear that $F \in \mathcal{C}'(A)$, and so it remains only to show that $\mathrm{d}(F^{-1}) : B \to \mathcal{L}(\mathbb{R}^n)$ is continuous. By Proposition 10.40, the bijection $F^{-1} : B \to A$ is continuous. Also, for each $\mathbf{x} \in A$ we found that $\mathrm{d}F_{\mathbf{x}}$ is invertible, and so the continuous mapping $\mathrm{d}F : A \to \mathcal{L}(\mathbb{R}^n)$ in fact maps A into $\mathrm{GL}_n(\mathbb{R})$. Finally, the map $\mathrm{Inv} : \mathrm{GL}_n(\mathbb{R}) \to \mathrm{GL}_n(\mathbb{R})$ given by $\mathrm{Inv}(L) = L^{-1}$ is continous by Theorem 10.34(3). Hence $\mathrm{Inv} \circ \mathrm{d}F \circ F^{-1} : B \to \mathrm{GL}_n(\mathbb{R})$ is continouous. Let $\mathbf{y} \in B$, so there is a unique $\mathbf{x} \in A$ such that $F(\mathbf{x}) = \mathbf{y}$. By (10.21),

$$(\mathrm{Inv} \circ \mathrm{d}F \circ F^{-1})(\mathbf{y}) = \mathrm{Inv}\big(\mathrm{d}F(\mathbf{x})\big) = \mathrm{Inv}(\mathrm{d}F_{\mathbf{x}}) = \mathrm{d}F_{\mathbf{x}}^{-1} = \mathrm{d}(F^{-1})_{\mathbf{y}} = \mathrm{d}(F^{-1})(\mathbf{y}),$$

so $\mathrm{d}(F^{-1}) = \mathrm{Inv} \circ \mathrm{d}F \circ F^{-1}$ and hence $\mathrm{d}(F^{-1})$ is continuous. Therefore $F^{-1} \in \mathcal{C}'(B)$. ∎

In the statement of the Inverse Function Theorem it is clear that the open set B is in fact $F(A)$; that is, A and $F(A)$ are both open in $\mathbb{R}^n$. This leads to another important result.

Theorem 10.56 (**Open Mapping Theorem**). *Let $U \subseteq \mathbb{R}^n$ be open, and suppose the mapping $F : U \to \mathbb{R}^n$ is such that $F \in \mathcal{C}'(U)$ and $\mathrm{d}F_{\mathbf{x}}$ is invertible for all $\mathbf{x} \in U$. Then $F(W)$ is open in $\mathbb{R}^n$ for all open $W \subseteq U$.*

Proof. Let $W \subseteq U$ be an open set. Fix $\mathbf{y} \in F(W)$, so there exists some $\mathbf{x} \in W$ such that $F(\mathbf{x}) = \mathbf{y}$. Since the mapping $F : W \to \mathbb{R}^n$ is such that $F \in \mathcal{C}'(W)$, and $\mathrm{d}F_{\mathbf{x}}$ is invertible, by the Inverse Function Theorem there exists an open set $A \subseteq W$ such that $\mathbf{x} \in A$ and $F(A)$ is open. Now, $\mathbf{y} = F(\mathbf{x}) \in F(A) \subseteq F(W)$, and so $\mathbf{y}$ is an interior point of $F(W)$. Since $\mathbf{y} \in F(W)$ is arbitrary, we conclude that $F(W)$ is open in $\mathbb{R}^n$. ■

In general, given metric spaces (X, d) and (Y, ρ), a function $f : (X, d) \to (Y, \rho)$ is an **open mapping** if $f(U)$ is open in (Y, ρ) whenever U is open in (X, d). Recalling the definition of the Jacobian of a mapping $F : S \subseteq \mathbb{R}^n \to \mathbb{R}^n$ given by (10.17), and also recalling (see the remark at the end of §10.2) that a linear operator $L \in \mathcal{L}(V)$ is invertible if and only if $\det(L) \neq 0$, the Open Mapping Theorem immediately implies the following.

Corollary 10.57. *Let $U \subseteq \mathbb{R}^n$ be open, and suppose $F : U \to \mathbb{R}^n$ is such that $F \in \mathcal{C}'(U)$. If $\det(\mathrm{d}F_{\mathbf{x}}) \neq 0$ for all $\mathbf{x} \in U$, then F is an open mapping.*

10.7 – The Implicit Function Theorem and Rank Theorem

We continue to regard the elements of any euclidean space to be column vectors, so that they interact in a natural way with matrices and conform to the usual format of systems of equations. If

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m,$$

then we define the vector

$$(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^{n+m}.$$

In general, whenever given that $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m}$, we assume $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.

For any $L \in \mathcal{L}(\mathbb{R}^{n+m}, \mathbb{R}^n)$, define $L_{\bullet 0} \in \mathcal{L}(\mathbb{R}^n)$ and $L_{0\bullet} \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ by

$$L_{\bullet 0}(\mathbf{x}) = L(\mathbf{x}, \mathbf{0}) \quad \text{and} \quad L_{0\bullet}(\mathbf{y}) = L(\mathbf{0}, \mathbf{y})$$

for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Then, for any $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m}$,

$$L(\mathbf{x}, \mathbf{y}) = L\big((\mathbf{x}, \mathbf{0}) + (\mathbf{0}, \mathbf{y})\big) = L(\mathbf{x}, \mathbf{0}) + L(\mathbf{0}, \mathbf{y}) = L_{\bullet 0}(\mathbf{x}) + L_{0\bullet}(\mathbf{y}). \tag{10.26}$$

If $L_{\bullet 0}$ or $L_{0\bullet}$ are invertible, then the symbols $L_{\bullet 0}^{-1}$ and $L_{0\bullet}^{-1}$ will denote their inverses. That is,

$$L_{\bullet 0}^{-1} := (L_{\bullet 0})^{-1} \quad \text{and} \quad L_{0\bullet}^{-1} := (L_{0\bullet})^{-1}.$$

Proposition 10.58. *Let $L \in \mathcal{L}(\mathbb{R}^{n+m}, \mathbb{R}^n)$. If $L_{\bullet 0}$ is invertible, then for each $\mathbf{k} \in \mathbb{R}^m$ there exists a unique $\mathbf{h} \in \mathbb{R}^n$ such that $L(\mathbf{h}, \mathbf{k}) = \mathbf{0}$. Moreover,*

$$\mathbf{h} = -(L_{\bullet 0}^{-1} \circ L_{0\bullet})(\mathbf{k}). \tag{10.27}$$

Proof. Suppose $L_{\bullet 0}$ is invertible, and fix $\mathbf{k} \in \mathbb{R}^m$. Applying (10.26),

$$\begin{aligned} L(\mathbf{h}, \mathbf{k}) &= L\big(-(L_{\bullet 0}^{-1} \circ L_{0\bullet})(\mathbf{k}), \mathbf{k}\big) = L_{\bullet 0}\big(-(L_{\bullet 0}^{-1} \circ L_{0\bullet})(\mathbf{k})\big) + L_{0\bullet}(\mathbf{k}) \\ &= -L_{\bullet 0}\big(L_{\bullet 0}^{-1}(L_{0\bullet}(\mathbf{k}))\big) + L_{0\bullet}(\mathbf{k}) = -L_{0\bullet}(\mathbf{k}) + L_{0\bullet}(\mathbf{k}) = \mathbf{0}, \end{aligned}$$

which proves existence and confirms (10.27).

Now suppose that $\mathbf{h} \in \mathbb{R}^n$ is such that $L(\mathbf{h}, \mathbf{k}) = \mathbf{0}$. Then $L_{\bullet 0}(\mathbf{h}) + L_{0\bullet}(\mathbf{k}) = \mathbf{0}$ by (10.26), and

$$L_{\bullet 0}(\mathbf{h}) + L_{0\bullet}(\mathbf{k}) = \mathbf{0} \;\Rightarrow\; L_{\bullet 0}(\mathbf{h}) = -L_{0\bullet}(\mathbf{k}) \;\Rightarrow\; \mathbf{h} = -L_{\bullet 0}^{-1}\big(L_{0\bullet}(\mathbf{k})\big)$$

since $L_{\bullet 0}$ is invertible. This gives (10.27), proving uniqueness. ∎

Proposition 10.59. *Let $U \subseteq \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$ and $G : U \to \mathbb{R}^\ell$ be such that $F, G \in \mathcal{C}'(U)$. Suppose $\Phi : U \to \mathbb{R}^{m+\ell}$ is given by*

$$H(\mathbf{x}) = \big(F(\mathbf{x}), G(\mathbf{x})\big)$$

for all $\mathbf{x} \in U$. Then $H \in \mathcal{C}'(U)$, and for each $\mathbf{a} \in U$,

$$\mathrm{d}H_{\mathbf{a}}(\mathbf{x}) = \big(\mathrm{d}F_{\mathbf{a}}(\mathbf{x}), \mathrm{d}G_{\mathbf{a}}(\mathbf{x})\big)$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. By Theorem 10.51, the partial derivatives $\partial_j F_i$ and $\partial_j G_k$ exist and are continuous on U for all $1 \leq i \leq m$, $1 \leq j \leq n$, and $1 \leq k \leq \ell$. Since the components of Φ are

$$\Phi_1 = F_1, \ldots, \Phi_m = F_m, \Phi_{m+1} = G_1, \ldots, \Phi_{m+\ell} = G_\ell,$$

we see that $\partial_j \Phi_i$ exist and are continuous on U for all $1 \leq i \leq m + \ell$ and $1 \leq j \leq n$. Therefore $\Phi \in \mathcal{C}'(U)$ by Theorem 10.51.

Next, fix $\mathbf{a} \in U$. Let

$$\alpha = \lim_{\mathbf{h} \to \mathbf{0}} \frac{H(\mathbf{a}+\mathbf{h}) - H(\mathbf{a}) - (\mathrm{d}F_{\mathbf{a}}(\mathbf{h}), \mathrm{d}G_{\mathbf{a}}(\mathbf{h}))}{\|\mathbf{h}\|}.$$

Then

$$\begin{aligned}
\alpha &= \lim_{\mathbf{h} \to \mathbf{0}} \frac{\big(F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a}) - \mathrm{d}F_{\mathbf{a}}(\mathbf{h}), G(\mathbf{a}+\mathbf{h}) - G(\mathbf{a}) - \mathrm{d}G_{\mathbf{a}}(\mathbf{h})\big)}{\|\mathbf{h}\|} \\
&= \lim_{\mathbf{h} \to \mathbf{0}} \left(\frac{F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a}) - \mathrm{d}F_{\mathbf{a}}(\mathbf{h})}{\|\mathbf{h}\|}, \frac{G(\mathbf{a}+\mathbf{h}) - G(\mathbf{a}) - \mathrm{d}G_{\mathbf{a}}(\mathbf{h})}{\|\mathbf{h}\|} \right) \\
&= (\mathbf{0}, \mathbf{0}),
\end{aligned}$$

and therefore

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|H(\mathbf{a}+\mathbf{h}) - H(\mathbf{a}) - \mathrm{d}H_{\mathbf{a}}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$$

for $\mathrm{d}H_{\mathbf{a}} = (\mathrm{d}F_{\mathbf{a}}, \mathrm{d}G_{\mathbf{a}})$, as was to be shown. ∎

Theorem 10.60 (Implicit Function Theorem). *Let $\Omega \subseteq \mathbb{R}^{n+m}$ be open, and let $F : \Omega \to \mathbb{R}^n$ be such that $F \in \mathcal{C}'(\Omega)$ and $F(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ for some $(\mathbf{a}, \mathbf{b}) \in \Omega$. If $\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,, \mathbf{0}) \in \mathcal{L}(\mathbb{R}^n)$ is invertible, then there exist open sets $U \subseteq \Omega$ and $W \subseteq \mathbb{R}^m$, with $(\mathbf{a}, \mathbf{b}) \in U$ and $\mathbf{b} \in W$, such that the following hold:*

1. *There is a unique function $G : W \to \mathbb{R}^n$ with $G(\mathbf{b}) = \mathbf{a}$, and such that*

$$(G(\mathbf{y}), \mathbf{y}) \in U \quad \textit{and} \quad F(G(\mathbf{y}), \mathbf{y}) = \mathbf{0}$$

   for all $\mathbf{y} \in W$.
2. *$G \in \mathcal{C}'(W)$ and*

$$\mathrm{d}G_{\mathbf{b}} = -\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,, \mathbf{0})^{-1} \circ \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0}, \,\cdot\,). \tag{10.28}$$

Proof. Suppose $dF_{(\mathbf{a},\mathbf{b})}(\cdot,\mathbf{0})$ is invertible. The mapping $\Lambda:\mathbb{R}^{n+m}\to\mathbb{R}^m$ given by $\Lambda(\mathbf{x},\mathbf{y})=\mathbf{y}$ is a linear mapping, so $\Lambda\in\mathcal{C}'(\mathbb{R}^{n+m})$ by Proposition 10.39, which implies that $\Lambda:\Omega\to\mathbb{R}^m$ is a $\mathcal{C}'$-mapping on Ω. Define $\Phi:\Omega\to\mathbb{R}^{n+m}$ by

$$\Phi(\mathbf{x},\mathbf{y})=\big(F(\mathbf{x},\mathbf{y}),\Lambda(\mathbf{x},\mathbf{y})\big)=\big(F(\mathbf{x},\mathbf{y}),\mathbf{y}\big).$$

Then $\Phi\in\mathcal{C}'(\Omega)$ by Proposition 10.59, and in particular Φ is differentiable at $(\mathbf{a},\mathbf{b})$. By Proposition 10.59 once more, $d\Phi_{(\mathbf{a},\mathbf{b})}:\mathbb{R}^{n+m}\to\mathbb{R}^{n+m}$ is given by

$$d\Phi_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{y})=\big(dF_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{y}),\mathbf{y}\big),$$

where $d\Lambda_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{y})=\Lambda(\mathbf{x},\mathbf{y})=\mathbf{y}$ by Proposition 10.39. Now, if $d\Phi_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{y})=(\mathbf{0},\mathbf{0})$, it follows that $dF_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{y})=\mathbf{0}$ and $\mathbf{y}=\mathbf{0}$, and hence $dF_{(\mathbf{a},\mathbf{b})}(\mathbf{x},\mathbf{0})=\mathbf{0}$. Since $dF_{(\mathbf{a},\mathbf{b})}(\cdot,\mathbf{0})$ is invertible, Theorem 10.26 and Proposition 10.21 imply that $\operatorname{Nul}(dF_{(\mathbf{a},\mathbf{b})}(\cdot,\mathbf{0}))=\{\mathbf{0}\}$, and so $\mathbf{x}=\mathbf{0}$. Thus $\operatorname{Nul}(d\Phi_{(\mathbf{a},\mathbf{b})})=\{(\mathbf{0},\mathbf{0})\}$, whereupon Proposition 10.21 implies $d\Phi_{(\mathbf{a},\mathbf{b})}$ is bijective and hence invertible.

We now apply the Inverse Function Theorem to $\Phi:\Omega\to\mathbb{R}^{n+m}$ to conclude that there exist open sets $U\subseteq\Omega$, $V\subseteq\Phi(U)$ such that $(\mathbf{a},\mathbf{b})\in U$, $(\mathbf{0},\mathbf{b})=\Phi(\mathbf{a},\mathbf{b})\in V$, and $\Phi:U\to V$ is a $\mathcal{C}'$-diffeomorphism. Let

$$W=\{\mathbf{y}\in\mathbb{R}^m:(\mathbf{0},\mathbf{y})\in V\},$$

so $W\subseteq\mathbb{R}^m$ is an open set with $\mathbf{b}\in W$.

Fix $\mathbf{y}\in W$. Then $(\mathbf{0},\mathbf{y})\in V$, and since $\Phi(U)=V$ there exists some $(\mathbf{x},\mathbf{y})\in U$ such that $\Phi(\mathbf{x},\mathbf{y})=(\mathbf{0},\mathbf{y})$, and therefore $F(\mathbf{x},\mathbf{y})=0$. If we suppose $\boldsymbol{\xi}\in\mathbb{R}^n$ is such that $(\boldsymbol{\xi},\mathbf{y})\in U$ and $F(\boldsymbol{\xi},\mathbf{y})=0$, then we obtain

$$\Phi(\boldsymbol{\xi},\mathbf{y})=\big(F(\boldsymbol{\xi},\mathbf{y}),\mathbf{y}\big)=\big(F(\mathbf{x},\mathbf{y}),\mathbf{y}\big)=\Phi(\mathbf{x},\mathbf{y}),$$

and thus $\boldsymbol{\xi}=\mathbf{x}$ since $\Phi:U\to V$ is injective. Therefore there is a unique function $G:W\to\mathbb{R}^n$ for which $(G(\mathbf{y}),\mathbf{y})\in U$ and $F(G(\mathbf{y}),\mathbf{y})=\mathbf{0}$ for all $\mathbf{y}\in W$. Moreover, since for $\mathbf{b}\in W$ we have $(\mathbf{a},\mathbf{b})\in U$ and $F(\mathbf{a},\mathbf{b})=\mathbf{0}$, it is clear that $G(\mathbf{a})=\mathbf{b}$.

Because $\Phi:U\to V$ is a $\mathcal{C}'$-diffeomorphism, the mapping $\Phi^{-1}:V\to U$ is a bijection such that $\Phi^{-1}\in\mathcal{C}'(V)$. Now, $(G(\mathbf{y}),\mathbf{y})\in U$ with $\Phi(G(\mathbf{y}),\mathbf{y})=(\mathbf{0},\mathbf{y})$ for all $\mathbf{y}\in W$, and so

$$\Phi^{-1}(\mathbf{0},\mathbf{y})=\big(G(\mathbf{y}),\mathbf{y}\big)$$

for all $\mathbf{y}\in W$. By Theorem 10.51 all partial derivatives of $\Phi^{-1}:V\to U$ are continuous on V, which implies that all partial derivatives of $\Phi^{-1}(\mathbf{0},\cdot):W\to U$ are continuous on W. Since the components of $G:W\to\mathbb{R}^n$ are also components of $\Phi^{-1}(\mathbf{0},\cdot)$, it follows that the partial derivatives of G must be continuous on W, and hence $G\in\mathcal{C}'(W)$ by Theorem 10.51.

It remains to verify (10.28). For brevity let $H=\Phi^{-1}(\mathbf{0},\cdot)$, and note that $H\in\mathcal{C}'(W)$. For any $\mathbf{y}\in W$ we have, by Proposition 10.59,

$$dH_{\mathbf{y}}(\mathbf{k})=\big(dG_{\mathbf{y}}(\mathbf{k}),\mathbf{k}\big)\tag{10.29}$$

for all $\mathbf{k}\in\mathbb{R}^m$. Also, since $H:W\to U$ and $F:U\to\mathbb{R}^n$ are each differentiable, and $F\circ H\equiv O_W$ on W, the Chain Rule gives

$$O=d(O_W)_{\mathbf{y}}=d(F\circ H)_{\mathbf{y}}=dF_{H(\mathbf{y})}\circ dH_{\mathbf{y}}$$

for all $\mathbf{y} \in W$. In particular

$$\mathrm{d}F_{(\mathbf{a},\mathbf{b})} \circ \mathrm{d}H_{\mathbf{b}} = O, \tag{10.30}$$

since $\mathbf{b} \in W$ with $H(\mathbf{b}) = (\mathbf{a}, \mathbf{b})$. Now, for any $\mathbf{k} \in \mathbb{R}^m$ we have

$$\begin{aligned}\big(\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0}) \circ \mathrm{d}G_{\mathbf{b}} + \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,)\big)(\mathbf{k}) &= \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0})(\mathrm{d}G_{\mathbf{b}}(\mathbf{k})) + \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,)(\mathbf{k}) \\ &= \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathrm{d}G_{\mathbf{b}}(\mathbf{k}), \mathbf{k})\end{aligned}$$

by (10.26), and then

$$\big(\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0}) \circ \mathrm{d}G_{\mathbf{b}} + \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,)\big)(\mathbf{k}) = \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathrm{d}H_{\mathbf{b}}(\mathbf{k})) = \mathbf{0}$$

by (10.29) and (10.30), respectively. Hence

$$\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0}) \circ \mathrm{d}G_{\mathbf{b}} + \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,) = O,$$

and since $\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0})$ is invertible, we finally obtain (10.28). ■

The following variant of the Implicit Function Theorem can be proven either in the same fashion as Theorem 10.60, or by reordering variables and applying Theorem 10.60 directly.

Corollary 10.61. *Let $\Omega \subseteq \mathbb{R}^{n+m}$ be open, and let $F : \Omega \to \mathbb{R}^m$ be such that $F \in \mathcal{C}'(\Omega)$ and $F(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ for some $(\mathbf{a}, \mathbf{b}) \in \Omega$. If $\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,) \in \mathcal{L}(\mathbb{R}^m)$ is invertible, then there exist open sets $U \subseteq \Omega$ and $W \subseteq \mathbb{R}^n$, with $(\mathbf{a}, \mathbf{b}) \in U$ and $\mathbf{a} \in W$, such that the following hold:*

1. *There is a unique function $G : W \to \mathbb{R}^m$ with $G(\mathbf{a}) = \mathbf{b}$, and such that*

$$(\mathbf{x}, G(\mathbf{x})) \in U \quad \text{and} \quad F(\mathbf{x}, G(\mathbf{x})) = \mathbf{0}$$

   for all $\mathbf{x} \in W$.
2. *$G \in \mathcal{C}'(W)$ and*

$$\mathrm{d}G_{\mathbf{a}} = -\mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\mathbf{0},\,\cdot\,)^{-1} \circ \mathrm{d}F_{(\mathbf{a},\mathbf{b})}(\,\cdot\,,\mathbf{0}).$$

For the next example we make extensive use of the definition

$$(x_1, \ldots, x_n) := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

for $x_1, \ldots, x_n \in \mathbb{R}$.

Example 10.62. Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable functions such that $f(1) = g(1) = 0$. Find conditions on the functions f and g which will permit solving the system of equations

$$\begin{cases} f(xy) + g(yz) = 0 \\ g(xy) + f(yz) = 0 \end{cases}$$

for y and z as functions of x in a neighborhood of the point $(1, 1, 1)$.

Solution. Define $\Phi : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$\Phi(x,y,z) = \big(f(xy) + g(yz), g(xy) + f(yz)\big).$$

The component functions

$$\varphi_1(x,y,z) = f(xy) + g(yz) \quad \text{and} \quad \varphi_2(x,y,z) = g(xy) + f(yz),$$

being continuously differentiable, have continuous first partials on $\mathbb{R}^3$ and therefore $\Phi \in \mathcal{C}'(\mathbb{R}^3)$. Also we have

$$\Phi(1,1,1) = \big(f(1) + g(1), g(1) + f(1)\big) = (0,0).$$

Now, for any point $\mathbf{a} = (a,b,c) \in \mathbb{R}^3$, the linear mapping $\mathrm{d}\Phi_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^2$ has matrix representation

$$[\mathrm{d}\Phi_{\mathbf{a}}] = \begin{bmatrix} \partial_x\varphi_1(\mathbf{a}) & \partial_y\varphi_1(\mathbf{a}) & \partial_z\varphi_1(\mathbf{a}) \\ \partial_x\varphi_2(\mathbf{a}) & \partial_y\varphi_2(\mathbf{a}) & \partial_z\varphi_2(\mathbf{a}) \end{bmatrix}.$$

Next, define the linear mapping $P : \mathbb{R}^2 \to \mathbb{R}^3$ by $P(y,z) = (0,y,z)$, which has matrix representation

$$[P] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

If $L = \mathrm{d}\Phi_{\mathbf{a}} \circ P$, then

$$L(y,z) = \mathrm{d}\Phi_{\mathbf{a}}(P(y,z)) = \mathrm{d}\Phi_{\mathbf{a}}(0,y,z)$$

and we see that $L = \mathrm{d}\Phi_{\mathbf{a}}(0,\cdot) \in \mathcal{L}(\mathbb{R}^2)$. By a proposition in §4.7 of [LAN] the matrix representation for L is

$$[L] = [\mathrm{d}\Phi_{\mathbf{a}}][P] = \begin{bmatrix} \partial_x\varphi_1(\mathbf{a}) & \partial_y\varphi_1(\mathbf{a}) & \partial_z\varphi_1(\mathbf{a}) \\ \partial_x\varphi_2(\mathbf{a}) & \partial_y\varphi_2(\mathbf{a}) & \partial_z\varphi_2(\mathbf{a}) \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \partial_y\varphi_1(\mathbf{a}) & \partial_z\varphi_1(\mathbf{a}) \\ \partial_y\varphi_2(\mathbf{a}) & \partial_z\varphi_2(\mathbf{a}) \end{bmatrix}.$$

In particular

$$[\mathrm{d}\Phi_{(1,1,1)}(0,\cdot)] = \begin{bmatrix} \partial_y\varphi_1(1,1,1) & \partial_z\varphi_1(1,1,1) \\ \partial_y\varphi_2(1,1,1) & \partial_z\varphi_2(1,1,1) \end{bmatrix},$$

and since

$$\partial_y\varphi_1(x,y,z) = \frac{\partial}{\partial y}\big[f(xy) + g(yz)\big] = xf'(xy) + zg'(yz),$$

$$\partial_z\varphi_1(x,y,z) = \frac{\partial}{\partial z}\big[f(xy) + g(yz)\big] = yg'(yz),$$

$$\partial_y\varphi_2(x,y,z) = \frac{\partial}{\partial y}\big[f(yz) + g(xy)\big] = zf'(yz) + xg'(xy),$$

and

$$\partial_z\varphi_2(x,y,z) = \frac{\partial}{\partial z}\big[f(yz) + g(xy)\big] = yf'(yz),$$

it follows that

$$[\mathrm{d}\Phi_{(1,1,1)}(0,\cdot)] = \begin{bmatrix} f'(1) + g'(1) & g'(1) \\ f'(1) + g'(1) & f'(1) \end{bmatrix}.$$

The linear mapping $d\Phi_{(1,1,1)}(0,\cdot)$ is invertible if and only if

$$\det\big([d\Phi_{(1,1,1)}(0,\cdot)]\big) = \begin{vmatrix} f'(1)+g'(1) & g'(1) \\ f'(1)+g'(1) & f'(1) \end{vmatrix} = [f'(1)]^2 - [g'(1)]^2 \neq 0;$$

that is, $d\Phi_{(1,1,1)}(0,\cdot)$ is invertible if and only if $f'(1) \neq \pm g'(1)$. If $f'(1) \neq \pm g'(1)$, then the Implicit Function Theorem implies there are open sets $U \subseteq \mathbb{R}^3$ and $I \subseteq \mathbb{R}$, with $(1,1,1) \in U$ and $1 \in I$, for which there is a (unique) function $\Psi : I \to \mathbb{R}^2$ with $\Psi(1) = (1,1)$ that satisfies

$$(x, \Psi(x)) \in U \quad \text{and} \quad \Phi(x, \Psi(x)) = (0,0)$$

for all $x \in I$. Letting $\psi_1, \psi_2 : I \to \mathbb{R}$ be the components of Ψ, we may write

$$\Phi(x, \psi_1(x), \psi_2(x)) = (0,0);$$

that is, for all $x \in I$, we have $y = \psi_1(x)$ and $z = \psi_2(x)$ such that

$$\big(f(xy) + g(yz), g(xy) + f(yz)\big) = (0,0),$$

and therefore

$$\begin{cases} f\big(x\psi_1(x)\big) + g\big(\psi_1(x)\psi_2(x)\big) = 0 \\ g\big(x\psi_1(x)\big) + f\big(\psi_1(x)\psi_2(x)\big) = 0 \end{cases}$$

Note that $(x, \psi_1(x), \psi_2(x)) \in U$ for all $x \in I$, where U is an open set containing $(1,1,1)$ as desired. ∎

If V and W are vector spaces, and $L \in \mathcal{L}(V,W)$, then the **rank** of L is defined to be

$$\operatorname{rank}(L) = \dim(\operatorname{Img}(L)).$$

A bijective linear mapping is called an **isomorphism**. Theorem 10.26 and Proposition 10.28 make it clear that every isomorphism has an inverse, and that inverse is also an isomorphism.

If V is a vector space, then a linear operator $\Pi \in \mathcal{L}(V)$ is a **projection in *V*** if $\Pi \circ \Pi = \Pi$. Thus, for all $\mathbf{v} \in V$, $\Pi(\Pi(\mathbf{v})) = \Pi(\mathbf{v})$.

Proposition 10.63. *Let V be a vector space.*

1. *If Π is a projection in V, then for every $\mathbf{v} \in V$ there exist unique vectors $\mathbf{u} \in \operatorname{Img}(\Pi)$ and $\mathbf{w} \in \operatorname{Nul}(\Pi)$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$.*
2. *If V is finite-dimensional and W is a subspace, then there exists a projection Π in V such that $\operatorname{Img}(\Pi) = W$.*

Proof.

Proof of Part (1). Suppose Π is a projection in V, and let $\mathbf{v} \in V$. Now,

$$\Pi(\mathbf{v}) = \Pi(\Pi(\mathbf{v})) \quad \Rightarrow \quad \Pi(\mathbf{v}) - \Pi(\Pi(\mathbf{v})) = \mathbf{0} \quad \Rightarrow \quad \Pi(\mathbf{v} - \Pi(\mathbf{v})) = \mathbf{0},$$

and so $\mathbf{v} - \Pi(\mathbf{v}) \in \operatorname{Nul}(\Pi)$. Let $\mathbf{u} = \Pi(\mathbf{v})$ and $\mathbf{w} = \mathbf{v} - \Pi(\mathbf{v})$. Then $\mathbf{u} \in \operatorname{Img}(\Pi)$ and $\mathbf{w} \in \operatorname{Nul}(\Pi)$ are such that $\mathbf{u} + \mathbf{w} = \mathbf{v}$, which proves existence.

Next, suppose $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$ for some $\mathbf{u}' \in \operatorname{Img}(\Pi)$ and $\mathbf{w}' \in \operatorname{Nul}(\Pi)$. Then $\Pi(\mathbf{w}) = \mathbf{0}$, and there exists some $\mathbf{z} \in V$ such that $\Pi(\mathbf{z}) = \mathbf{u}'$. Now,

$$\Pi(\mathbf{z}) = \mathbf{u}' \quad \Rightarrow \quad \Pi(\Pi(\mathbf{z})) = \Pi(\mathbf{u}') \quad \Rightarrow \quad \Pi(\mathbf{z}) = \Pi(\mathbf{u}') \quad \Rightarrow \quad \mathbf{u}' = \Pi(\mathbf{u}'),$$

so that

$$\mathbf{u}' = \mathbf{u}' + \mathbf{0} = \Pi(\mathbf{u}') + \Pi(\mathbf{w}') = \Pi(\mathbf{u}' + \mathbf{w}') = \Pi(\mathbf{v}) = \mathbf{u},$$

and then

$$\mathbf{w}' = \mathbf{v} - \mathbf{u}' = \mathbf{v} - \Pi(\mathbf{v}) = \mathbf{w}.$$

Thus $\mathbf{u}' = \mathbf{u}$ and $\mathbf{w}' = \mathbf{w}$, which proves uniqueness.

Proof of Part (2). Suppose V is finite-dimensional and W is a subspace. By Theorem 10.17(2)

$$m = \dim(W) \leq \dim(V) = n$$

for some $m, n \in \mathbb{W}$. If either $m = 0$ or $n = 0$, then we may let $\Pi = O_V$ and the proof is done. Thus we may henceforth assume that $m, n \in \mathbb{N}$.

Let $\mathcal{B}_W = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a basis for W. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent in V, and so by Theorem 10.16 vectors $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ may be found such that $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V. Define $\Pi \in \mathcal{L}(V)$ as follows:

$$\Pi(\mathbf{v}_k) = \begin{cases} \mathbf{v}_k, & 1 \leq k \leq m \\ \mathbf{0}, & m+1 \leq k \leq n. \end{cases}$$

Fix $\mathbf{v} \in V$, so that $\mathbf{v} = \sum_{k=1}^{n} a_k \mathbf{v}_k$ for some $a_1, \ldots, a_n \in \mathbb{R}$. Then by linearity and Theorem 10.4,

$$\Pi(\mathbf{v}) = \sum_{k=1}^{n} a_k \Pi(\mathbf{v}_k) = \sum_{k=1}^{m} a_k \mathbf{v}_k \in W,$$

which shows that $\operatorname{Img}(\Pi) \subseteq W$. Moreover,

$$\Pi(\Pi(\mathbf{v})) = \Pi\left(\sum_{k=1}^{m} a_k \mathbf{v}_k\right) = \sum_{k=1}^{m} a_k \Pi(\mathbf{v}_k) = \sum_{k=1}^{m} a_k \mathbf{v}_k = \Pi(\mathbf{v}),$$

which shows that $\Pi \circ \Pi = \Pi$. Finally, if $\mathbf{w} \in W$ so that $\mathbf{w} = \sum_{k=1}^{m} b_k \mathbf{v}_k$ for some $b_1, \ldots, b_m \in \mathbb{R}$, we obtain $\Pi(\mathbf{w}) = \mathbf{w}$, whence $\mathbf{w} \in \operatorname{Img}(\Pi)$, and therefore $\operatorname{Img}(\Pi) = W$. ■

It is a fact that any subspace of $\mathbb{R}^n$ is a closed set. For instance, in $\mathbb{R}^3$ the only possible subspaces are $\{\mathbf{0}\}$, lines through $\mathbf{0}$, planes through $\mathbf{0}$, and $\mathbb{R}^3$ itself. Thus, in the statement of the following lemma is must be kept in mind that if V is a subspace of $\mathbb{R}^n$ and $A \subseteq V$ is given to be open in V, so that $A = V \cap B$ for some set B open in $\mathbb{R}^n$, then A is not necessarily open in $\mathbb{R}^n$ unless $V = \mathbb{R}^n$.

Lemma 10.64. *Let V, W be vector subspaces of $\mathbb{R}^n, \mathbb{R}^m$, respectively, with each given the standard subspace topology. If $L \in \mathcal{L}(V, W)$ is surjective, then L is an open mapping.*

The following is essentially the Rank Theorem as stated in [Rud]. Aside from some remarks that come after, the proof is omitted.

Theorem 10.65 (Rank Theorem). *Suppose $m, n, r \in \mathbb{W}$ with $m, n \geq r$, let $W \subseteq \mathbb{R}^n$ be open, and suppose $F : W \to \mathbb{R}^m$ is a $\mathcal{C}'$-mapping with $\operatorname{rank}(\mathrm{d}F_{\mathbf{x}}) = r$ for all $\mathbf{x} \in W$. For fixed $\mathbf{a} \in W$ let $\Pi \in \mathcal{L}(\mathbb{R}^m)$ be a projection such that $\Pi(\mathbb{R}^m) = \mathrm{d}F_{\mathbf{a}}(\mathbb{R}^n)$. Then there exist open $U, V \subseteq \mathbb{R}^n$ with $\mathbf{a} \in U \subseteq W$, a $\mathcal{C}'$-diffeomorphism $H : V \to U$, and a $\mathcal{C}'$-mapping $\varphi : \mathrm{d}F_{\mathbf{a}}(V) \to \operatorname{Nul}(\Pi)$, such that*

$$(F \circ H)(\mathbf{x}) = \mathrm{d}F_{\mathbf{a}}(\mathbf{x}) + \varphi(\mathrm{d}F_{\mathbf{a}}(\mathbf{x}))$$

for all $\mathbf{x} \in V$.

First note that $\mathrm{d}F_{\mathbf{a}} : \mathbb{R}^n \to \mathrm{d}F_{\mathbf{a}}(\mathbb{R}^n)$ is a surjective linear mapping, where of course $\mathrm{d}F_{\mathbf{a}}(\mathbb{R}^n)$ is a subspace of $\mathbb{R}^m$. Since V is open in $\mathbb{R}^n$, by Lemma 10.64 it follows that $\mathrm{d}F_{\mathbf{a}}(V)$ is open in $\mathrm{d}F_{\mathbf{a}}(\mathbb{R}^n)$. This does not mean that $\mathrm{d}F_{\mathbf{a}}(V)$ is open in $\mathbb{R}^m$, however. Thus, for the statement of the theorem to make sense, we must generalize Definition 10.49 modestly to include sets that are not necessarily open in a euclidean space, but rather open in a *subspace* of the euclidean space.[4]

[4]This is not done in [Rud], and so what it means for the function φ in the Rank Theorem to be a $\mathcal{C}'$-mapping on $\mathrm{d}F_{\mathbf{a}}(V)$ is left a dangling ambiguity. Ideally we would here employ the machinery of charts, atlases, and coordinate maps from the theory of smooth manifolds. This will be done in a separate document sometime in the future.

10.9 – Derivatives of Higher Order

10.10 – Differentiation of Integrals

Theorem 10.66 (**Leibniz's Rule**). *Let* $\varphi : [a,b] \times [c,d] \to \mathbb{R}$ *be continuous, and define* $g : [c,d] \to \mathbb{R}$ *by*

$$g(t) = \int_a^b \varphi(x,t)\,dx.$$

Then g *is continuous. If in addition* $\partial_t \varphi$ *exists and is continuous on* $[a,b] \times [c,d]$, *then* g *is continuously differentiable with*

$$g'(t) = \int_a^b \partial_t \varphi(x,t)\,dx$$

for each $t \in [c,d]$.

Proof. Fix $t_0 \in [c,d]$, and let $\epsilon > 0$. Since φ is uniformly continuous on $R := [a,b] \times [c,d]$, there exists $\delta > 0$ such that

$$\|\mathbf{x} - \mathbf{y}\| < \delta \;\;\Rightarrow\;\; \big|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\big| < \frac{\epsilon}{b-a}$$

for all $\mathbf{x}, \mathbf{y} \in R$. ■