Differentiation in Banach Spaces

Robert L. Bocchino Jr.

Revised January 2026

This paper is a supplement to my papers *The General Derivative* and *The Inverse and Implicit Mapping Theorems*. To keep the presentation simple, those papers focused on finite-dimensional vector spaces. This paper extends the theory to general Banach spaces, including spaces of infinite dimension (e.g., function spaces).

To read §§ 1–6 of this paper, you should be familiar with the material presented in *The General Derivative*. To read § 7, you should also be familiar with the material presented in *The Inverse and Implicit Mapping Theorems*.

1. Fields

We begin by defining the concept of a field, which generalizes the real numbers.

A **field** F is a set of elements satisfying the following rules:

1. **Addition:** We can add any two elements a and b of F to form the element $a + b$ of F. Addition is associative and commutative. It has a **zero element** 0 such that for any element a of F, $a + 0 = 0 + a = a$. For every element a of F there is an element of F called the **additive inverse** of a and written $-a$ such that $a + (-a) = (-a) + a = 0$.
2. **Multiplication:** We can multiply any two elements a and b of F to form the element $a \cdot b$ or ab of F. Multiplication is associative and commutative, and it distributes over addition. F has a **multiplicative identity** 1 such that for any element a of F, $1 \cdot a = a \cdot 1 = a$. For every element a of F except 0, there is an element of F called the **multiplicative inverse** of a and written a^{-1} such that $aa^{-1} = a^{-1}a = 1$.

Note that these are exactly the rules of addition and multiplication t we stated for the real numbers **R** in § 1 of *The General Derivative*. By the same argument we gave in *The General Derivative*, for real numbers, we can establish that $0a = a0 = 0$ for any element a of F.

Let F be a field. A **norm** or **absolute value** on F is a function that assigns to every element a of F a real number $|a|$, such that the following rules are satisfied:

1. For any element a of F, $|a| \geq 0$, and $|a| = 0$ if and only if $a = 0$.
2. For any two elements a_1 and a_2 of F, $|a_1 a_2| = |a_1||a_2|$.
3. For any two elements a_1 and a_2 of F, $|a_1 + a_2| \leq |a_1| + |a_2|$.

The real number $|a|$ is called the **norm** of a. The last inequality is called the **triangle inequality**. Note that these are exactly the rules for the norm or absolute value of a real number r that we stated in § 1 of *The General Derivative*. Note also that in the case of the real norm, r and $|r|$ are both real numbers, while in the case of a field norm, a is an element of F (not necessarily **R**), and $|a|$ is a real number.

A field with a norm is called a **normed field**. In the theory of differentiation, the most common normed fields are the real numbers **R** and the complex numbers **C** with their usual absolute values. The rational numbers **Q** with the usual absolute value also constitute a normed field.

2. Vector Spaces

Next we define the concept of a vector space in the general setting. Let F be a field. A **vector space** V over F is a set of **vectors** satisfying the following rules.

1. **Vector addition:** We can add any two vectors v_1 and v_2 to form the vector $v_1 + v_2$. Addition is associative and commutative, it has a zero element 0, and each vector v has an additive inverse $-v$ such that $v + (-v) = -v + v = 0$.

2. **Scalar multiplication:** The elements of F are called the **scalars** of V. Given any scalar a and any vector v, we can form the product av. We also write va, and this has the same meaning as av. This product satisfies the following rules:
   a. For all vectors v, $1v = v$.
   b. Scalar multiplication is associative. That is, if a_1 and a_2 are scalars and v is a vector, then $a_1(a_2v) = (a_1a_2)v$.
   c. Scalar multiplication distributes over addition. That is, if a_1 and a_2 are scalars and v_1 and v_2 are vectors, then $(a_1 + a_2)v = a_1v + a_2v$ and $a(v_1 + v_2) = av_1 + av_2$.

This is exactly the definition of a vector space that we gave in *The General Derivative*, after replacing **R** with F in the definition. By the same argument as for real vector spaces, we can establish that $0v = 0$ for all vectors v.

Normed vector spaces: Let F be a normed field (§ 1), and let V be a vector space over F. A **norm** on V is a function that assigns to each vector of v a real number $|v|$, such that the following rules are satisfied:

1. For any vector v, $|v| \geq 0$, and $|v| = 0$ if and only if $v = 0$.
2. For any scalar a and vector v, $|av| = |a||v|$.
3. For any two vectors v_1 and v_2, $|v_1 + v_2| \leq |v_1| + |v_2|$.

The number $|v|$ is called the **norm** of v. Again this is exactly the definition of a norm that we gave in *The General Derivative*, after replacing **R** with F in the definition.

A vector space V with a norm is called a **normed vector space**. For example, the real numbers **R**, the complex numbers **C**, and the rational numbers **Q** are normed vector spaces. In practice, most applications of the theory of differentiation occur in normed vector spaces over **R** or **C**.[1]

Topological vector spaces: A normed vector space is an example of a more general construct called a **topological vector space**. A topological vector space V is a vector space with a **topology**. A topology designates certain subsets of V as **open sets** in a way that satisfies certain rules. In a normed vector space, a set S is open if and only if, for every vector v in S, there is a real number $\varepsilon > 0$ such that all vectors v' with $|v - v'| < \varepsilon$ lie in S. This definition satisfies the rules for a topology, so it makes V into a topological vector space.

Other kinds of topologies are possible. In general, a topological vector space is a vector space V over a field F together with a topology satisfying the following rules:

TVS1 For every scalar a in F, the function $M(a)\colon V \to V$ given by $M(a)(x) = ax$ is continuous.

TVS2 For every vector v in V, the function $A(v)\colon V \to V$ given by $A(v)(x) = v + x$ is continuous.

A map $f\colon X \to Y$ of topological spaces is continuous if, for every open set $U \subseteq Y$, the set $f^{-1}(U)$ of elements x in X such that $f(x) \in U$ is an open set. Note that this definition does not depend on the existence of a norm. It is easy to show that, for a normed vector space, this definition of continuity is equivalent to the one given in § 3 of *The General Derivative*. It is also easy to show that the topology of a normed vector space satisfies rules **TVS1** and **TVS2**.

3. Completeness

We define maps, limits, and continuity for normed vector spaces exactly as in § 3 of *The General Derivative*. In the more general setting, we also need to define the notion of completeness. We do that in this section.

Sequences of vectors: Let V be a vector space. A **sequence** of vectors in V is an enumeration of vectors, one for each natural number $i = 0, 1, \ldots$. We use the notation $\{v_i\}_{i \in \mathbf{N}}$ to denote a sequence. The subscript i called the **index** of the element v_i of the sequence. We may also write a sequence by listing its initial elements, if the elements repeat after a certain index, or if the elements follow a repeating pattern.

For example, consider the following sequences of real numbers:

1. $\{1\}_{i \in \mathbf{N}}$ denotes the sequence $1, 1, 1, \ldots$ that assigns the number 1 at every index i.

[1] Some authors say that a normed vector space must be over **R** or **C**. Others say that a normed vector space must be over a subfield of **C**; this definition includes **R**, **C**, and **Q**. Here we adopt the most general definition: any normed field is allowed. We just have to be mindful that an arbitrary normed field F may behave differently from **R**, **C**, or **Q**. For example, F may not be complete as a vector space over itself (§ 3), or it may be be non-Archimedean (see https://en.wikipedia.org/wiki/Non-Archimedean_ordered_field). When we need the properties of a specific scalar field such as **R** or **C**, we will specify it.

2. $\{1/2^{-i}\}_{i \in \mathbf{N}}$ denotes the sequence $1, 1/2, 1/4, \ldots$.
3. $\{(-1)^i\}_{i \in \mathbf{N}}$ denotes the sequence $1, -1, 1, -1, \ldots$.
4. For all i in **N**, let π_i denote the real number formed by multiplying π by 10^i, taking the integer part of the result, and dividing by 10^i. Then $\{\pi_i\}_{i \in \mathbf{N}}$ denotes the sequence 3, 3.1, 3.14, 3.141, $\ldots$ formed by adding successive digits in the decimal representation of π.

Convergent sequences: Let V be a normed vector space, and let $S = \{v_i\}_{i \in \mathbf{N}}$ be a sequence. Informally, we say that S **converges** to a vector v if the elements of S become arbitrarily close to v once the indices i become large enough.[2] Formally, for any $\varepsilon > 0$, there exists some $N \geq 0$ such that $|v_i - v| < \varepsilon$ for all $i \geq N$. If a sequence S converges to some vector v, then we say that S **converges** or **is convergent**.

For example:

1. $\{1\}_{i \in \mathbf{N}}$ converges to 1.
2. $\{1/2^{-i}\}_{i \in \mathbf{N}}$ converges to 0.
3. $\{(-1)^i\}_{i \in \mathbf{N}}$ does not converge.
4. $\{\pi_i\}_{i \in \mathbf{N}}$ converges to π.

Cauchy sequences: Let $S = \{v_i\}_{i \in \mathbf{N}}$ be a sequence. Informally, we say that S is **Cauchy** if the elements of S become arbitrarily close to each other once we go far enough out in the sequence. Formally, for any $\varepsilon > 0$ and any natural number N, we say that S is **ε-Cauchy at** N if for every i and j such that $i \geq N$ and $j \geq N$, $|v_i - v_j| < \varepsilon$. We say that S is Cauchy if, for any $\varepsilon > 0$, there is a natural number N such that S is ε-Cauchy at N.

If a sequence is convergent, then it is Cauchy. Indeed, suppose a sequence $S = \{v_i\}_{i \in \mathbf{N}}$ converges to v. Then for any ε, there is a natural number N such that for all $i \geq N$ and $j \geq N$, we have $|v - v_i| < \varepsilon/2$ and $|v - v_j| < \varepsilon/2$. Therefore

$$|v_i - v_j| = |v_i - v + v - v_j| \leq |v_i - v| + |v - v_j| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so S is ε-Cauchy at N. In the third step, we have used the triangle inequality.

Complete normed vector spaces: Let V be a normed vector space over a field F. We say that V is **complete** if every Cauchy sequence in V converges. For example, **R** is complete as a normed vector space over itself, but **Q** is not. To see that **Q** is not complete, consider the sequence $S = \{\pi_i\}_{i \in \mathbf{N}}$. As a sequence in **R**, S is Cauchy and converges to π. As a sequence in **Q**, S is Cauchy but does not converge to any element of **Q**.

Another normed vector space that is not complete is $P[0,1]$, the vector space over **R** of polynomial functions with coefficients in **R** defined on the closed interval $[0,1] \subseteq \mathbf{R}$. There are several norms one can put on $P[0,1]$. For example, for any polynomial $p(x)$ in $P[0,1]$, we can let $|p(x)|$ be the maximum value of $|p(a)|$ such that a lies in $[0,1]$. Then we can use the Taylor series expansion of e^x in $[0,1]$ to construct a sequence of functions in $P[0,1]$ that converge to the function e^x on $[0,1]$, which is not a polynomial function.

A complete normed vector space is called a **Banach space**. Every finite-dimensional normed vector space over **R** or **C** is a Banach space. This is why, in *The General Derivative*, we didn't have to define or specify the property of completeness. We worked in finite dimensions over **R**, so we got completeness "for free."

An example of an infinite-dimensional Banach space over **R** is the space of bounded continuous functions from **R** to **R** with the sup norm. This is the same as Example 5 from § 2 of *The General Derivative*, with the additional requirement that the functions f are continuous.

Complete topological vector spaces: As with continuity, one can generalize the concept of completeness so it doesn't depend on having a norm. One can then define the concept of a complete topological vector space, and this generalizes the concept of a Banach space. An important example is a **Fréchet space**. This is a complete topological vector space whose topology satisfies some further conditions.[3]

4. Linear Maps

Bounded linear maps: In § 4.2 of *The General Derivative*, we defined the norm $|\lambda|$ of a linear map λ from a finite-dimensional normed vector space V over **R** to a finite-dimensional normed vector space W over **R**. We said that $|\lambda|$

[2] The phrase "arbitrarily close" is a bit of mathematics jargon. It means "as close as desired" or "close to within any specified tolerance."

[3] See https://en.wikipedia.org/wiki/Fr%C3%A9chet_space.

is the supremum of all values $|\lambda(v)|$ such that $|v| \leq 1$. Because λ is linear, we could also take the norm to be the supremum of all values $|\lambda(v)|$ such that $|v| = 1$; these definitions are equivalent. In finite dimensions over **R**, the norm $|\lambda|$ is well-defined for any linear map λ, because the supremum is guaranteed to exist.

In general, when V and W are normed vector spaces over a field F, the norm of a linear map λ as defined in the previous paragraph is not guaranteed to exist. When it does exist, then we say that λ is a **bounded linear map**.

Continuous linear maps: A bounded linear map is continuous. Indeed, let V and W be normed vector spaces over a field F, and let $\lambda\colon V \to W$ be a bounded linear map. From § 4.2 of *The General Derivative*, we have

$$|\lambda(x)| \leq |\lambda||x| \tag{1}$$

for all vectors x. Suppose we want to show that λ is continuous at v. Given $\varepsilon > 0$, let $\delta = \varepsilon/|\lambda|$. Then by (1), for all x such that $|x - v| < \delta$, we have

$$|\lambda(x - v)| \leq |\lambda||x - v| < |\lambda|\delta = |\lambda|(\varepsilon/|\lambda|) = \varepsilon,$$

which was to be shown.

When F is **Q**, **R**, or **C**, a continuous linear map is bounded. Indeed, let V and W be normed vector spaces over a field F, where F is **Q**, **R**, or **C**, and let $\lambda\colon V \to W$ be a continuous linear map. In particular, λ is continuous at zero, so we can choose a rational number $\delta > 0$ such that for all x such that $|x| < \delta$, we have $|\lambda(x)| < 1$. Then for all x such that $|x| < 2$, we have $|x\delta/2| < \delta$, so

$$|\lambda(x\delta/2)| = |(\delta/2)\lambda(x)| = |\delta/2||\lambda(x)| < 1.$$

Therefore $|\lambda(x)| < 2/\delta$ when $|x| \leq 1$, so λ is bounded.

Thus when we are working over **Q**, **R**, or **C**, a linear map is bounded if and only if it is continuous. Moreover, the concept of a continuous linear map does not depend on the existence of a norm, so we can apply it to topological vector spaces such as Fréchet spaces that don't have norms.

Isomorphisms: Let V and W be topological vector spaces over a field F. An **isomorphism** from V to W is a continuous linear map $\lambda\colon V \to W$ satisfying the following conditions:

1. λ is **bijective**, or one-to-one. That is, for every vector w in W, there is one and only one vector v in V such that $\lambda(v) = w$. Therefore λ has an **inverse function** $\lambda^{-1}\colon W \to V$. This is the function that takes each vector w in W to the vector v specified above. When a linear map is bijective, we also say that it is **invertible**.
2. λ^{-1} is a continuous linear map.

When an isomorphism $\lambda\colon V \to W$ exists, then we say that V and W are **isomorphic**, and we write $V \cong W$.

In a Banach space or Fréchet space over **R** or **C**, if a continuous linear map λ is invertible, then it is an isomorphism. This result follows from the **open mapping theorem**, which says that in a Banach space or Fréchet space over **R** or **C**, the image $\lambda(U)$ of an open set U under a continuous linear map λ is an open set. Here $\lambda(U)$ is the set of all elements $\lambda(v)$ such that v is an element of U. The open mapping theorem is not true for general topological vector spaces.

In the rest of this document, we will work in Banach spaces over **R** or **C**. In these spaces, a linear map is bounded if and only if it is continuous, and a bounded (or continuous) linear map is an isomorphism if and only if it is invertible. In the special case of finite dimensions over **R** or **C**, every normed vector space is a Banach space, and every linear map is continuous, so every invertible linear map is an isomorphism.

The vector space $L(V, W)$: Let V and W be Banach spaces over F, where F is **R** or **C**. We define $L(V, W)$ to be the space of continuous linear maps $\lambda\colon V \to W$, with the norm $|\lambda|$. Because each map λ is continuous and therefore bounded, $|\lambda|$ is well-defined. This definition generalizes the definition we gave in § 4.2 of *The General Derivative* for the normed vector space $L(V, W)$ in finite dimensions over **R**.

It is straightforward to show that $L(V, W)$ is complete, and therefore a Banach space. Here is a sketch of the proof:

1. Let $S = \{\lambda_i\}_{i \in \mathbf{N}}$ be any Cauchy sequence of bounded linear maps $\lambda_i\colon V \to W$. We must show that S converges to a bounded linear map $\lambda\colon V \to W$.
2. Use the completeness of W to show that for any vector v in V, the sequence $\{\lambda_i(v)\}_{i \in \mathbf{N}}$ converges to an element of w. Construct a function $\lambda\colon V \to W$ by setting $\lambda(v) = w$.

3. Show that λ is a bounded linear map.
   a. Use the linearity of each λ_i and the properties of limits in W to show that λ is a linear map.
   b. Observe that since S is Cauchy, the sequence $\{|\lambda_i|\}$ is bounded by a constant M. Show that M is a bound for λ.
4. Observe that since S is Cauchy, for any $\varepsilon > 0$ we can fix an $N \geq 0$ such that for all $i \geq N$ and $j \geq N$, $|\lambda_i - \lambda_j| < \varepsilon$. Then for any x in V with $|x| \leq 1$, we have $|\lambda_i(x) - \lambda_j(x)| < \varepsilon$. Therefore $|\lambda_i(x) - \lambda(x)| < \varepsilon$ for all $|x| \leq 1$ and all $i \geq N$, so $|\lambda_i - \lambda| < \varepsilon$ for all $i \geq N$. Therefore S converges to λ in $L(V, W)$.

5. The Derivative

With the theory of Banach spaces in hand, we can generalize the definition of the derivative that we gave in *The General Derivative*. Let X and Y be Banach spaces over F, where F is **R** or **C**. Let $U \subseteq X$ be a set, let $f : U \to Y$ be a map, and let x be a point of X. We say that f is **differentiable** at x if there is an open neighborhood H of zero such that $x + h \in U$ for all $h \in H$, a bounded linear map $Df(x) : X \to Y$, and a function $\phi(h) : H \to Y$ that is $o(h)$ such that

$$f(x + h) = f(x) + Df(x)(h) + \phi(h)$$

for all $h \in H$. We say that the linear map $Df(x)$ is the **deriviative** of f at x. This is the same definition we gave in § 5 of *The General Derivative*, except that we are more precise about the domains of f and ϕ, and we explicitly require the linear map $Df(x)$ to be bounded.

With this definition, all of the theory presented in §§ 5–9 of *The General Derivative* for finite-dimensional vector spaces over **R** goes through for Banach spaces over **R** or **C**.

6. Splitting of Banach Spaces

In this section we discuss the decomposition or "splitting" of Banach spaces into products of subspaces. This kind of decomposition is useful in applications such as differential geometry.

An example: We begin with an example. Consider the vector space $V = \mathbf{R}^3$ with the Euclidean norm. Let V_1 be the set of elements $(x, 0, 0)$, in $\mathbf{R}^3$ such that $x \in \mathbf{R}$. With the obvious rules for addition and scalar multiplication, i.e.,

$$(v, 0, 0) + (v', 0, 0) = (v + v', 0, 0)$$

$$r \cdot (v, 0, 0) = (rv, 0, 0),$$

V_1 is a vector space over **R**. It is isomorphic to **R**. It is a **subspace** of $\mathbf{R}^3$, i.e., all of its elements are also elements of $\mathbf{R}^3$.[4] Moreover, it is what we call a **closed subspace** of $\mathbf{R}^3$. In topology, a closed set is a set S whose complement (i.e., the set of all points not in S) is open. Here the set $\mathbf{R}^3 - V_1$ (i.e., the set of all points in $\mathbf{R}^3$ that are not in V_1) is open, so V_1 is closed.

Now let V_2 be the set of elements $(0, x_1, x_2)$ of $\mathbf{R}^3$ such that x_1 and x_2 are real numbers. V_2 is a closed subspace of $\mathbf{R}^3$ that is isomorphic to $\mathbf{R}^2$. Observe the following about V_1 and V_2:

1. V_1 and V_2 are closed subspaces of $\mathbf{R}^3$.
2. The intersection of V_1 and V_2 contains one element $(0, 0, 0)$, the zero element of the addition law of $\mathbf{R}^3$.
3. Every element in $\mathbf{R}^3$ may be represented as a sum of vectors $v_1 + v_2$, with v_1 in V_1 and v_2 in V_2.
4. The representation in element 3 is an isomorphism from $V_1 \times V_2$ (with the sup norm) to $\mathbf{R}^3$.

We summarize items 2 and 3 above by saying that V_1 and V_2 are **complementary** subspaces of $\mathbf{R}^3$. We summarize items 1 through 4 above by saying that $\mathbf{R}^3$ **splits** into the subspaces V_1 and V_2.

The general definition: Let V be a topological vector space, and let V_1 and V_2 be complementary closed subspaces of V. If the map $\phi : V_1 \times V_2 \to V$ given by $\phi(v_1, v_2) = v_1 + v_2$ is an isomorphism (i.e., a continuous linear map with a

[4] Notice that if we are being precise, then we do not say that **R** itself is a subspace of $\mathbf{R}^3$. The elements of **R** are real numbers, and the elements of $\mathbf{R}^3$ are triples of real numbers. Also, $\mathbf{R}^3$ has many subspaces that are isomorphic to **R**, not just the one we chose. For example, we could have chosen the set of elements $(0, x, 0)$ such that $x \in \mathbf{R}$. In practice mathematicians often do say that **R** is a subspace of $\mathbf{R}^3$. What they mean is that it is possible to construct an isomorphism from **R** to a subspace of $\mathbf{R}^3$. Such an isomorphism from a space to a subset of a larger space is sometimes called an **embedding**.

continuous linear inverse), then we say that V **splits** into the subspaces V_1 and V_2. Notice how this definition generalizes the example we gave above of splitting $\mathbf{R}^3$ into subspaces isomorphic to $\mathbf{R}^1$ and $\mathbf{R}^2$.

Now let V be a Banach space over $\mathbf{R}$ or $\mathbf{C}$. In this case, from the theory of topological vector spaces, we know the following:

1. If V_1 and V_2 are closed complementary subspaces of V, then V splits into V_1 and V_2.
2. If V_1 is a finite-dimensional closed subspace of V, then there exists a closed complementary subspace V_2. Therefore V splits into V_1 and V_2.

See [Lang 2001], § 2. Notice that these statements are true in the special case that $V = \mathbf{R}^n$, $V_1 \cong \mathbf{R}^{m_1}$, $V_2 \cong \mathbf{R}^{m_2}$, and $m_1 + m_2 = n$.

Coordinate systems: In § 6 of *The General Derivative*, we showed how to compute derivatives over vector spaces $V = V_1 \times \cdots \times V_n$ that are products of other vector spaces. We called such a product a **coordinate system**. When a vector space V splits into subspaces V_1 and V_2, then by definition it is isomorphic to the product $V_1 \times V_2$. Therefore we say that it splits into the coordinate system $V_1 \times V_2$.

7. The Inverse and Implicit Mapping Theorems

We now revisit the results presented in *The Inverse and Implicit Mapping Theorems*. We extend the results to Banach spaces and develop some useful applications.

7.1. The Inverse Mapping Theorem

In this section, we let X and Y be Banach spaces over F, where F is $\mathbf{R}$ or $\mathbf{C}$. We let $U \subseteq X$ be an open set and $f\colon U \to Y$ be a map that is continuously differentiable to order $n > 0$.

Theorem (Inverse Mapping). Assume that $Df(x)$ is invertible at each point $x \in U$. Then at each point $p \in U$, f has a local inverse, i.e., there exists an open neighborhood $V \subseteq U$ of p and an invertible map $g\colon V \to f(V)$ such that $g = f$ on V. Moreover, g^{-1} is continuously differentiable to order n, and at each point y in $f(V)$ we have $Dg^{-1}(y) = Df(g^{-1}(y))^{-1}$.

The proof is as given in § 1.5 of *The Inverse and Implicit Mapping Theorems*. Note the following:

1. $Df(x)$ is a bounded linear map by definition (§ 5).
2. In a Banach space, an invertible bounded linear map is an isomorphism, i.e., its inverse is also bounded (§ 4). So in the statement of the theorem, we could also have said, "Assume that $Df(x)$ is an isomorphism at each point $x \in U$."

Using the theory of infinite series, one can show that the set of invertible bounded linear maps from X to Y is open in $L(X, Y)$.[5] Then, using the fact that $Df(x)$ is continuous, one can show that if $Df(p)$ is invertible at some point $p \in U$, then there exists an open neighborhood $W \subseteq U$ of p such that the conditions of the theorem are satisfied at the points of W. Therefore we have the following corollary:

Corollary. Assume that $Df(p)$ is invertible for some point $p \in U$. Then f has a local inverse at p, i.e., there exists an open neighborhood $V \subseteq U$ of p and an invertible map $g\colon V \to f(V)$ such that $g = f$ on V. Moreover, g^{-1} is continuously differentiable to order n, and $Dg^{-1}(f(p)) = Df(p)^{-1}$.

Proof: By the comments above, we can choose an open neighborhood $W \subseteq U$ of p and apply the theorem. □

When Y is a coordinate system: Assume that Y splits into a coordinate system $Y_1 \times Y_2$ (§ 6), and let $f_1\colon U \to Y_1$ and $f_2\colon U \to Y_2$ be the coordinate maps of f. Fix a point p in U. Assume that $Df_1(p)\colon X \to Y_1$ is an isomorphism, and $Df_2(p)\colon X \to Y_2$ is the zero map. Equivalently, $Df(p)\colon X \to Y$ is an injection (so it is isomorphic with its image $Df(p)(X)$), $Df(p)(X)$ is a closed subspace of Y, and Y splits into $Df(p)(X)$ and a complementary closed subspace. In this case, we may let Y_1 be $Df(p)(X)$ and let Y_2 be the complementary closed subspace.

Under these conditions, in applications such as differential geometry, it is useful to construct an open neighborhood $V \subseteq Y$ of $f(p)$ and a map $g\colon V \to Y$ such that g is continuously differentiable to order n, g has a local inverse at $f(p)$ that is continuously differentiable to order n, and for all x in $f^{-1}(V)$,

$$g(f(x)) = (f_1(x), 0).$$

[5] See https://en.wikipedia.org/wiki/Neumann_series.

Notice that the obvious choice of $g(y_1, y_2) = (y_1, 0)$ does not satisfy the requirements. While this map produces the required values, it does not have a local inverse at $f(p)$.

By the corollary, there is an open neighborhood $W \subseteq U$ of p and a map $h: W \to f_1(W)$ such that $h = f_1$ on W, h has an inverse h^{-1}, and h^{-1} is continuously differentiable to order n. Because h^{-1} is continuous, $f_1(W) = h(W)$ is an open set. Let $V = h(W) \times Y_2$. Then $V \subseteq Y$ is an open neighborhood of $f(p)$. Now let $g: V \to Y$ be the map

$$g(y) = (g_1(y), g_2(y)) = (y_1, y_2 - f_2(h^{-1}(y_1)))$$

for any point $y = (y_1, y_2)$ in V. Then g is continuously differentiable to order n, since it is a composition of maps with that property. Further, we have the following partial derivatives at $f(p)$:

$$D_1 g_1(f(p)) = I \quad D_2 g_1(f(p)) = 0$$

$$D_1 g_2(f(p)) = Df_2(h^{-1}(f(p))) \circ D(h^{-1})(f(p)) = Df_2(p) \circ D(h^{-1})(f(p)) = 0 \circ D(h^{-1})(f(p)) = 0$$

$$D_2 g_2(f(p)) = I$$

Therefore $Dg(f(p)) = I$, so by the corollary, g^{-1} is continuously differentiable to order n. Finally, for any point x in $f^{-1}(V)$, we have

$$\begin{aligned} g(f(x))) &= (f_1(x), f_2(x) - f_2(h^{-1}(f_1(x)))) \\ &= (f_1(x), f_2(x) - f_2(x)) \\ &= (f_1(x), 0), \end{aligned}$$

as required.

When X is a coordinate system: Now assume that X splits into a coordinate system $X_1 \times X_2$, and let $U = U_1 \times U_2$. Fix a point $p = (p_1, p_2)$ in U, and assume that $D_2 f(p): X_2 \to Y$ is an isomorphism, where $D_2 f$ is the partial derivative of f with respect to X_2. Equivalently, $Df(p): X \to Y$ is a surjection, its kernel (i.e., the set of points x such that $Df(p)(x) = 0$) is a closed subspace of X, and X splits into the kernel of $Df(p)$ and a complementary closed subspace. In this case, we may let X_1 be the kernel and let X_2 be the complementary closed subspace.

Under these conditions, in applications such as differential geometry, is is useful to construct an open neighborhood $V \subseteq U$ of p and a map $g: V \to X$ such that g is continuously differentiable to order n, g has a local inverse at p that is continuously differentiable to order n, and for all $x = (x_1, x_2)$ in V,

$$f(g(x_1, x_2)) = f(p_1, x_2).$$

Notice that the obvious choice of $g(x_1, x_2) = (p_1, x_2)$ does not satisfy the requirements, because it does not have a local inverse at p.

As in § 6.2 of *The General Derivative*, write $f_{(p_1,-)}: U_2 \to Y$ to denote the map obtained from f by setting $x_1 = p_1$ and letting x_2 range over U_2. By definition, the partial derivative $D_2 f(p_1, x_2)$ is the derivative $Df_{(p_1,-)}(x_2)$. Since $D_2 f(p) = Df_{(p_1,-)}(p_2)$ is an isomorphism, by the corollary there is an open neighborhood $W \subseteq U_2$ of p_2 and a map $h: W \to f_{(p_1,-)}(W)$ such that $h = f_{(p_1,-)}$ on W, h has an inverse h^{-1}, and h^{-1} is continuously differentiable to order n. Because h^{-1} is continuous, $h(W)$ is an open set. Let $W' \subseteq X$ be the set $f^{-1}(h(W))$; by the continuity of f, W' is an open neighborhood of p. Now define $G: W' \to X$ as follows:

$$G(x_1, x_2) = (x_1, h^{-1}(f(x_1, x_2))).$$

By construction, G takes p to p and is continuously differentiable to order n. We will now show that (1) G has a local inverse g at p and (2) g has the required properties.

G has a local inverse g: We have the following partial derivatives:

$$D_1 G_1(p) = I \quad D_2 G_1(q) = 0$$

$$D_1 G_2(p) = D_1(h^{-1})(f(p)) \circ D_1 f(p)$$

$$D_2 G_2(p) = D_2(h^{-1})(f(p)) \circ D_2 f(p) = (D_2 f(p))^{-1} \circ D_2 f(p) = I$$

Therefore the Jacobian matrix for $DG(p)$ is

$$\begin{bmatrix} I & 0 \\ D_1(h^{-1})(f(p)) \circ D_1 f(p) & I \end{bmatrix}.$$

This matrix is invertible with inverse

$$\begin{bmatrix} I & 0 \\ -D_1(h^{-1})(f(p)) \circ D_1 f(p) & I \end{bmatrix}.$$

Therefore by the corollary, G has a local inverse $g\colon V \to X$, where V is an open neighborhood of p.

g has the required properties: By the corollary, g is continuously differentiable to order n. Further, from the definition of G, g has the following explicit formula:

$$\begin{aligned} g(x_1, x_2) &= (x_1, (f_{(x_1,-)})^{-1}(h(x_2))) \\ &= (x_1, (f_{(x_1,-)})^{-1}(f(p_1, x_2))) \end{aligned}$$

You can see this by using the definition of G to compute $g(G(x_1, x_2)) = (x_1, x_2)$. The inverse $(f_{(x_1,-)})^{-1}$ must exist because the function g exists. But then we have

$$\begin{aligned} f(g(x_1, x_2)) &= f(x_1, (f_{(x_1,-)})^{-1}(f(p_1, x_2)))) \\ &= f_{(x_1,-)}((f_{(x_1,-)})^{-1}(f(p_1, x_2))) \\ &= f(p_1, x_2), \end{aligned}$$

as required.

7.2. The Implicit Mapping Theorem

Theorem (Implicit Mapping). Let X_1, X_2, and Y be Banach spaces spaces over F, where F is **R** *or* **C**. *Let $U_1 \subseteq X_1$ and $U_2 \subseteq X_2$ be open sets, let $U = U_1 \times U_2$, and let $f\colon U \to Y$ be a map. Assume that f is continuously differentiable to order $n > 0$ and that the derivative $Df(x)$ is invertible at each point $x \in U$. Let $a = (a_1, a_2)$ be a point in U, and let $b = f(a)$. Then there exists an open neighborhood W_1 of a_1 in U_1 and a map $g\colon W_1 \to U_2$ such that $g(a_1) = a_2$, $f(x_1, g(x_1)) = b$ for all x_1 in W_1, and g is continuously differentiable to order n. Moreover, there exists a real number $r > 0$ such that the values $g(x)$ are uniquely determined for all $x \in B(a_1, r)$.*

The proof is as given in § 2.3 of *The Inverse and Implicit Mapping Theorems*.

References

Bocchino, R. *The General Derivative. https://rob-bocchino.net/Professional/Mathematics.html.*

Bocchino, R. *The Inverse and Implicit Mapping Theorems. https://rob-bocchino.net/Professional/Mathematics.html.*

Lang, Serge. *Fundamentals of Differential Geometry.* Springer Verlag 2001.

Lang, Serge. *Real and Functional Analysis.* Third Edition. Springer Verlag 1993.