

# Rudin: Functions of Several Variables

James Pagan

October 2023

## Contents

<b>1</b>	<b>Normed Vector Spaces</b>	<b>2</b>
1.1	Definition . . . . .	2
1.2	Matrix Norm . . . . .	3
1.3	Properties of Linear Maps . . . . .	4
1.4	Completeness of $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ . . . . .	5
<b>2</b>	<b>Differentiation</b>	<b>7</b>
2.1	The Derivative . . . . .	7
2.2	Chain Rule . . . . .	8
2.3	The Partial Derivative . . . . .	9
2.4	Mixed Partial Derivatives . . . . .	10
2.5	Real-Valued Functions . . . . .	11
<b>3</b>	<b>The Inverse Function Theorem</b>	<b>12</b>
3.1	The Contraction Principle . . . . .	12
3.2	The Inverse Function Theorem . . . . .	13
3.3	The Implicit Function Theorem . . . . .	15

# 1 Normed Vector Spaces

## 1.1 Definition

A **normed vector space** is a complex vector space  $X$  equipped with a mapping  $\|\cdot\| : X \rightarrow \mathbb{R}$  that satisfies the following properties:

1. **Positivity:**  $\|\mathbf{x}\| \geq 0$ , with equality if and only if  $\mathbf{x} = \mathbf{0}$ .
2. **Homogeneity:**  $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$  for all  $\lambda \in \mathbb{C}$ .
3. **Triangle Inequality:**  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

A complete normed vector space is called a **Banach space**. As demonstrated in RealAnalysis/babyrudin3.tex, all the desired limit formulas hold on normed vector spaces.

**Theorem 1.** *Norms are uniformly continuous mappings from  $X$  to  $\mathbb{R}$ .*

*Proof.* For all  $\epsilon > 0$ , we have that

$$0 < \|\mathbf{x} - \mathbf{y}\| < \epsilon \implies \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\| < \epsilon.$$

We conclude that  $\|\cdot\| : X \rightarrow \mathbb{R}$  is uniformly continuous.  $\square$

**Theorem 2.** *Let  $X$  be a finite-dimensional vector space over  $\mathbb{C}$ . Then any two norms on  $X$  are equivalent.*

*Proof.* Let  $\dim X = n$ . We first prove the theorem for  $\mathbb{C}^n$ ; let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be the canonical basis of  $\mathbb{C}^n$ , and suppose  $\|\cdot\|_1 : \mathbb{C}^n \rightarrow [0, \infty)$  is a norm. We prove that  $\|\mathbf{z}\|_1$  is equivalent to the canonical norm  $\|\mathbf{z}\|$ .

Consider the boundary of the unit ball of the canonical norm in  $\mathbb{C}^n$ . Since  $\|\cdot\|_1$  is continuous, the Extreme Value Theorem guarantees that there exists  $\mathbf{u}, \mathbf{s} \in X$  with canonical norm 1 such that

$$\|\mathbf{u}\|' = \inf_{\|\mathbf{z}\|=1} \{\|\mathbf{z}\|'\} \quad \text{and} \quad \|\mathbf{s}\|' = \sup_{\|\mathbf{z}\|=1} \{\|\mathbf{z}\|'\}$$

Then for all  $\mathbf{z} \in \mathbb{C}^n$ , the constants  $\|\mathbf{u}\|'$  and  $\|\mathbf{s}\|'$  allow for norm equivalence:

$$\|\mathbf{u}\|'\|\mathbf{z}\| \leq \left\| \frac{\mathbf{z}}{\|\mathbf{z}\|} \right\|' \|\mathbf{z}\| = \|\mathbf{z}\|' = \left\| \frac{\mathbf{z}}{\|\mathbf{z}\|} \right\|' \|\mathbf{z}\| \leq \|\mathbf{s}\|'\|\mathbf{z}\|.$$

We conclude that all norms on  $\mathbb{C}^n$  are equivalent to the canonical norm. Proving norm equivalence from  $X$  to  $\mathbb{C}^n$  is not challenging.  $\square$

## 1.2 Matrix Norm

Let  $X$  and  $Y$  be normed vector spaces. If  $\mathbf{T}$  is bounded, the **norm** of  $\mathbf{T}$  is as follows:

$$\|\mathbf{T}\| = \sup\{\|\mathbf{T}\mathbf{x}\| \mid \mathbf{x} \in X, \|\mathbf{x}\| \leq 1\}.$$

The **critical vector** of  $\mathbf{T}$  is the vector  $\mathbf{x} \in X$  such that  $\|\mathbf{x}\| \leq 1$  and  $\|\mathbf{T}\mathbf{x}\| = \|\mathbf{T}\|$ ; the critical vector always has norm 1. Naturally,  $\|\mathbf{T}\mathbf{x}\| \leq \|\mathbf{T}\| \cdot \|\mathbf{x}\|$ ; since equality is attained,  $\|\mathbf{T}\mathbf{x}\| \leq \lambda\mathbf{x}$  implies  $\|\mathbf{T}\| \leq \lambda$ .

**Theorem 3.** *If  $\mathbf{T}, \mathbf{S} \in \mathcal{B}(X, Y)$ , then  $\|\mathbf{T} + \mathbf{S}\| \leq \|\mathbf{T}\| + \|\mathbf{S}\|$ .  $X = Y$  entails  $\|\mathbf{TS}\| \leq \|\mathbf{T}\|\|\mathbf{S}\|$ .*

*Proof.* Let  $\mathbf{x}$  be the critical vector of  $\mathbf{T} + \mathbf{S}$ . Then

$$\|\mathbf{T} + \mathbf{S}\| = \|(\mathbf{T} + \mathbf{S})\mathbf{x}\| \leq \|\mathbf{T}\mathbf{x}\| + \|\mathbf{S}\mathbf{x}\| \leq \|\mathbf{T}\|\|\mathbf{x}\| + \|\mathbf{S}\|\|\mathbf{x}\| = \|\mathbf{T}\| + \|\mathbf{S}\|.$$

Now suppose  $X = Y$  and let  $\mathbf{w}$  be the critical vector of  $\mathbf{TS}$ . Then

$$\|\mathbf{TS}\| = \|\mathbf{TS}\mathbf{w}\| \leq \|\mathbf{T}\|\|\mathbf{S}\mathbf{w}\| \leq \|\mathbf{T}\|\|\mathbf{S}\|\|\mathbf{w}\| = \|\mathbf{T}\|\|\mathbf{S}\|.$$

This completes the proof. □

**Theorem 4.** *The matrix norm is a metric of all bounded linear maps in  $\mathcal{B}(X, Y)$ .*

*Proof.* Suppose  $\mathbf{T}, \mathbf{S} \in \mathcal{B}(X, Y)$  are bounded. We must perform four rather routine calculations:

1. **Positivity:** The matrix norm is nonnegative. If  $\|\mathbf{T} - \mathbf{S}\| = 0$ , then  $\|\mathbf{x}\| = 1$  implies  $(\mathbf{T} - \mathbf{S})\mathbf{x} = \mathbf{0}$ ; hence for all  $\mathbf{x} \in X$ ,

$$(\mathbf{T} - \mathbf{S})\mathbf{x} = \|\mathbf{x}\| \left( (\mathbf{T} - \mathbf{S}) \left( \frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right) = \|\mathbf{x}\|(0) = 0;$$

thus  $\mathbf{T} - \mathbf{S} = \mathbf{0}$  and  $\mathbf{T} = \mathbf{S}$ .

2. **Symmetry:** Notice that  $(\mathbf{T} - \mathbf{S})\mathbf{x} = -(\mathbf{S} - \mathbf{T})\mathbf{x}$  for all  $\mathbf{x} \in X$ . Naturally if  $\mathbf{w}$  is the critical vector of  $\mathbf{T} - \mathbf{S}$ , then  $-\mathbf{w}$  is the critical vector of  $\mathbf{S} - \mathbf{T}$ ; thus

$$\|\mathbf{T} - \mathbf{S}\| = \|\mathbf{w}\| = \|-\mathbf{w}\| = \|\mathbf{S} - \mathbf{T}\|.$$

3. **Triangle Inequality:** For all bounded  $\mathbf{R} \in \mathcal{L}(X, Y)$ ,

$$\|\mathbf{T} - \mathbf{S}\| = \|(\mathbf{T} - \mathbf{R}) + (\mathbf{R} - \mathbf{S})\| \leq \|\mathbf{T} - \mathbf{R}\| + \|\mathbf{R} - \mathbf{S}\|.$$

We conclude that the matrix norm is a metric of the bounded matrices of  $\mathcal{L}(X, Y)$ . □

It is straightforward that  $\|\lambda\mathbf{T}\| = |\lambda|\|\mathbf{T}\|$  for all  $\lambda \in \mathbb{C}$  as well.

### 1.3 Properties of Linear Maps

**Theorem 5.** *Let  $X$  be a finite-dimensional normed vector space, and let  $Y$  be a normed vector space. Then all  $\mathbf{T} \in \mathcal{L}(X, Y)$  are bounded.*

*Proof.* Let  $\dim X = n$  and let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be a basis of  $X$ . Then for all  $\mathbf{z} = z_1\mathbf{e}_1 + \dots + z_n\mathbf{e}_n$ , we have

$$\|\mathbf{Tz}\| \leq |z_1|\|\mathbf{Te}_1\| + \dots + |z_n|\|\mathbf{Te}_n\| \leq C(|z_1| + \dots + |z_n|),$$

where  $C = \max\{\|\mathbf{Te}_i\|\}$ . Realize that  $\|\mathbf{z}\|_1 = |z_1| + \dots + |z_n|$  defines a norm on  $X$ ; since all norms finite-dimensional vector spaces are equivalent, there exists another constant  $M$  such that  $|z_1| + \dots + |z_n| = \|\mathbf{z}\|_1 \leq M\|\mathbf{z}\|$ . Therefore

$$\|\mathbf{Tz}\| \leq CM\|\mathbf{z}\|,$$

so  $\mathbf{T}$  is bounded. This completes the proof.  $\square$

**Theorem 6.** *If  $\mathbf{T} \in \mathcal{L}(X, Y)$  is bounded, then  $\mathbf{T}$  is uniformly continuous.*

*Proof.* Let  $\mathbf{T}$  be bounded. For all  $\epsilon > 0$ , then  $0 \leq \|\mathbf{x} - \mathbf{y}\| < \frac{\epsilon}{\|\mathbf{T}\|}$  implies

$$\|\mathbf{T}\mathbf{x} - \mathbf{T}\mathbf{y}\| \leq \|\mathbf{T}\| \cdot \|\mathbf{x} - \mathbf{y}\| < \|\mathbf{T}\| \left( \frac{\epsilon}{\|\mathbf{T}\|} \right) = \epsilon.$$

Thus,  $\mathbf{T}$  is uniformly continuous.  $\square$

Let  $\Omega$  be the set of all bounded, invertible linear operators on  $X$ . Recall from Linear Algebra that an operator  $\mathbf{T} \in \mathcal{L}(\mathbb{C}^n)$  is invertible if and only if  $\text{range } \mathbf{T} = \mathbb{C}^n$ .

**Theorem 7.** *Let  $X$  be a finite dimensional vector space. If  $\mathbf{T} \in \Omega$  and  $\mathbf{S} \in \mathcal{L}(X)$  are both bounded, then*

$$\|\mathbf{T}^{-1}\| \cdot \|\mathbf{S} - \mathbf{T}\| < 1$$

*implies  $\mathbf{S} \in \Omega$ .*

*Proof.* Suppose  $\mathbf{S} \notin \Omega$ . Then there exists  $\mathbf{x} \in X$  of norm 1 such that  $\mathbf{S}\mathbf{x} = \mathbf{0}$ , so

$$\begin{aligned} 1 &= \|\mathbf{x}\| \\ &= \|\mathbf{T}^{-1}\mathbf{T}\mathbf{x}\| \\ &\leq \|\mathbf{T}^{-1}\| \cdot \|\mathbf{T}\mathbf{x}\| \\ &= \|\mathbf{T}^{-1}\| \cdot \|\mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{x}\| \\ &\leq \|\mathbf{T}^{-1}\| \cdot \|\mathbf{S} - \mathbf{T}\| \cdot \|\mathbf{x}\| \\ &= \|\mathbf{T}^{-1}\| \cdot \|\mathbf{S} - \mathbf{T}\|. \end{aligned}$$

Taking the contrapositive yields the desired result.  $\square$

**Theorem 8.** *Let  $X$  be a finite-dimensional normed vector space. Then  $\Omega$  is an open subset of  $\mathcal{L}(X)$ , and the bijection  $f : \mathbf{T} \rightarrow \mathbf{T}^{-1}$  is continuous on  $\Omega$ .*

*Proof.* Let  $\mathbf{T} \in \Omega$ . Since  $\|\mathbf{T}^{-1}\|$  is nonzero, we may consider the open ball at  $\mathbf{T}$  of radius  $\frac{1}{\|\mathbf{T}^{-1}\|}$ ; more specifically, all  $\mathbf{S} \in \mathcal{L}(\mathbb{C}^n)$  such that

$$\|\mathbf{T} - \mathbf{S}\| < \frac{1}{\|\mathbf{T}^{-1}\|}.$$

Since  $\|\mathbf{T}\|$  is nonzero,  $\|\mathbf{T} - \mathbf{S}\| \cdot \|\mathbf{T}^{-1}\| < 1$ ; hence  $\mathbf{S} \in \Omega$ . The open ball is contained within  $\Omega$ , so the set  $\Omega$  is open. As per continuity, realize that

$$\lim_{\mathbf{S} \rightarrow \mathbf{T}^{-1}} \|\mathbf{T} - \mathbf{S}\|$$

As per continuity: **Rudin's proof is nonrigorous**, and I don't know how to rectify it. The basic idea is that

$$\|\mathbf{T}^{-1} - \mathbf{S}^{-1}\| = \|\mathbf{T}^{-1}\mathbf{S}\mathbf{S}^{-1} - \mathbf{T}^{-1}\mathbf{T}\mathbf{S}^{-1}\| \leq \|\mathbf{T}^{-1}\| \cdot \|\mathbf{S} - \mathbf{T}\| \cdot \|\mathbf{S}^{-1}\|,$$

but using this to bound  $\epsilon$  depends on  $\mathbf{T}^{-1}$  and  $\mathbf{S}^{-1}$ . **I will leave this unfinished until instructor clarification.**  $\square$

## 1.4 Completeness of $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$

Before we discuss the completeness of  $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ , we must uncover an important inequality. Let  $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$ ; for any  $\mathbf{T} \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ , define the components of  $\mathbf{T}$  as  $t_{ij}$  for  $i \in \{1, \dots, n\}$  and  $j \in \{1, \dots, m\}$ . Then

$$\mathbf{T}\mathbf{z} = \left( \sum_{j=1}^n t_{1j}z_j \right) \mathbf{e}_1 + \cdots + \left( \sum_{j=1}^n t_{mj}z_j \right) \mathbf{e}_m.$$

Then via the Cauchy-Schwarz Inequality,

$$\begin{aligned} \|\mathbf{T}\mathbf{z}\|^2 &= \left( \sum_{j=1}^n t_{1j}z_j \right)^2 + \cdots + \left( \sum_{j=1}^n t_{mj}z_j \right)^2 \\ &\leq \left( \sum_{j=1}^n t_{1j}^2 \right) \left( \sum_{j=1}^n z_j^2 \right) + \cdots + \left( \sum_{j=1}^n t_{mj}^2 \right) \left( \sum_{j=1}^n z_j^2 \right) \\ &= \left( \sum_{j=1}^n z_j^2 \right) \left( \sum_{i=1}^m \sum_{j=1}^n t_{ij}^2 \right) = \|\mathbf{z}\|^2 \left( \sum_{i=1}^m \sum_{j=1}^n t_{ij}^2 \right). \end{aligned}$$

Let  $\mathbf{w}$  be the critical vector of  $\mathbf{T}$ . Then

$$\|\mathbf{T}\| = \|\mathbf{T}\mathbf{w}\| \leq \|\mathbf{z}\| \sqrt{\sum_{i=1}^m \sum_{j=1}^n t_{ij}^2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n t_{ij}^2}.$$

While  $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$  and  $\mathbb{C}^{n+m}$  may be isomorphic, the relevant bijection is *not* an isometry.

**Theorem 9.** *If  $Y$  is a Banach space, then  $\mathcal{B}(X, Y)$  is a Banach space.*

*Proof.* Let  $(\mathbf{T}_n)$  be a Cauchy sequence in  $\mathcal{B}(X, Y)$ ; for all  $\epsilon > 0$ , there exists  $N$  such that

$$N \leq n, m \implies \|\mathbf{T}_n - \mathbf{T}_m\| < \epsilon. \quad (1)$$

We will define the limit of  $(\mathbf{T}_n)$ . For any  $\mathbf{x} \in X$ , we have that

$$\|\mathbf{T}_n \mathbf{x} - \mathbf{T}_m \mathbf{x}\| \leq \|\mathbf{T}_n - \mathbf{T}_m\| \|\mathbf{x}\| \leq \epsilon \|\mathbf{x}\|.$$

By selecting  $\frac{\epsilon}{\|\mathbf{x}\|}$  in equation (1), we find that  $(\mathbf{T}_n \mathbf{x})$  is a Cauchy sequence in  $Y$ . Thus, it converges to a unique vector in  $Y$ . Define a mapping  $\mathbf{T} : X \rightarrow Y$  as follows:

$$\mathbf{T}\mathbf{z} \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \mathbf{T}_n \mathbf{z}$$

It is relatively easy to show that  $\mathbf{T}$  is linear: for all  $\mathbf{x}, \mathbf{y} \in X$  and  $\lambda \in \mathbb{C}$ , we have

$$\begin{aligned} \mathbf{T}(\mathbf{x} + \mathbf{y}) &= \lim_{n \rightarrow \infty} \mathbf{T}_n(\mathbf{x} + \mathbf{y}) = \lim_{n \rightarrow \infty} \mathbf{T}_n \mathbf{x} + \lim_{n \rightarrow \infty} \mathbf{T}_n \mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{T}\mathbf{y} \\ \mathbf{T}(\lambda \mathbf{x}) &= \lim_{n \rightarrow \infty} \mathbf{T}_n(\lambda \mathbf{x}) = \lambda \lim_{n \rightarrow \infty} \mathbf{T}_n \mathbf{x} = \lambda \mathbf{T}\mathbf{x}. \end{aligned}$$

We must show that  $\mathbf{T}$  is bounded and is the limit of  $(\mathbf{T}_n)$ . Observe that

$$\|\mathbf{T}\mathbf{z}\| \leq \|\mathbf{T}\mathbf{z} - \mathbf{T}_n \mathbf{z}\| + \|\mathbf{T}_n \mathbf{z}\|.$$

Observe that the transformation  $\phi : \mathbf{y} \rightarrow \|\mathbf{T}_n \mathbf{x} - \mathbf{y}\|$  is continuous, since it is a composition of continuous functions. Hence

$$\|\mathbf{T}_n \mathbf{z} \mathbf{x} - \mathbf{T}_m \mathbf{z}\| \leq \epsilon \|\mathbf{z}\| \implies \|\mathbf{T}_n - \mathbf{T}_m\| = \lim_{m \rightarrow \infty} \|\mathbf{T}_n \mathbf{z} - \mathbf{T}_m \mathbf{z}\| \leq \epsilon \|\mathbf{x}\|.$$

Then pick  $\epsilon = 1$  and  $n = N$ . We find that

$$\begin{aligned} \|\mathbf{T}\mathbf{x}\| &\leq \|(\mathbf{T} - \mathbf{T}_N)\mathbf{x}\| + \|\mathbf{T}_N \mathbf{x}\| \\ &\leq \|\mathbf{x}\| + \|\mathbf{T}_N\| \|\mathbf{x}\| \\ &\leq (1 + \|\mathbf{T}_N\|) \|\mathbf{x}\|. \end{aligned}$$

Letting  $c = 1 + \|\mathbf{T}_N\|$  yields that  $\mathbf{T}$  is bounded. As per the limit condition, we have that

$$\|\mathbf{T}_n - \mathbf{T}\| = \sup_{\mathbf{x} \neq 0} \frac{\|(\mathbf{T}_n - \mathbf{T})\mathbf{x}\|}{\|\mathbf{x}\|} \leq \epsilon,$$

which completes the proof. □

**Corollary 1.** *If the components of  $\mathbf{T} \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$  are continuous functions from a metric space  $X$  to  $\mathbb{R}$ , then the mapping  $X \rightarrow \mathbf{T}$  is continuous.*

*Proof.* Let the continuous components be  $f_{ij}$ . For all  $\epsilon > 0$ , there are  $N_{ij}$  such that

$$0 < d(x, y) < \delta_{ij} \implies |f_{ij}(x) - f_{ij}(y)| < \frac{\epsilon}{\sqrt{mn}}.$$

Then identical means as Theorem 6 demonstrate that the mapping  $X \rightarrow \mathbf{T}$  is continuous.  $\square$

## 2 Differentiation

### 2.1 The Derivative

Let  $f : E \rightarrow \mathbb{R}^m$  for an open set  $E \subset \mathbb{R}^n$ . Then  $f$  is **differentiable** at  $\mathbf{x} \in E$  if there exists a linear map  $D \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  such that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} = 0,$$

we say that  $f$  is **differentiable** at  $\mathbf{x}$  and write  $f'(\mathbf{x}) = \mathbf{J}$ , where  $J$  is the **total derivative** of  $f$  at  $\mathbf{x}$  — also called the matrix of partial derivatives, the differential, or the total derivative. If  $f$  is differentiable at *all*  $\mathbf{x} \in U$ , we say that  $f$  itself is differentiable over  $U$ .

**Lemma 1.** *The total derivative is unique.*

*Proof.* Define  $f$  like above. Suppose that for contradiction that there exist two matrices  $\mathbf{J} \neq K$  such that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - K\mathbf{h}\|}{\|\mathbf{h}\|} = 0.$$

See that  $\mathbf{J} - K \neq 0$ , so  $\|\mathbf{J} - K\| > 0$ . Then there exist  $d_1$  and  $d_2$  such that

$$\begin{aligned} 0 < \|\mathbf{h}\| < \delta_1 &\implies \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} < \frac{\|\mathbf{J} - K\|}{2} \\ 0 < \|\mathbf{h}\| < \delta_2 &\implies \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - K\mathbf{h}\|}{\|\mathbf{h}\|} < \frac{\|\mathbf{J} - K\|}{2} \end{aligned}$$

For  $0 < \|\mathbf{h}\| < \min\{\delta_1, \delta_2\}$ , we have that

$$\begin{aligned}\|\mathbf{J} - K\| &= \frac{\|J - K\|}{2} + \frac{\|J - K\|}{2} \\ &> \frac{\| -f(\mathbf{x} + \mathbf{h}) + f(\mathbf{x}) + \mathbf{J}\mathbf{h} \|}{\|\mathbf{h}\|} + \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - K\mathbf{h}\|}{\|\mathbf{h}\|} \\ &\geq \frac{\|(-f(\mathbf{x} + \mathbf{h}) + f(\mathbf{x}) + \mathbf{J}\mathbf{h}) + (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - K\mathbf{h})\|}{\|\mathbf{h}\|} \\ &= \frac{\|(\mathbf{J} - K)\mathbf{h}\|}{\|\mathbf{h}\|},\end{aligned}$$

so  $\|\mathbf{J} - K\|\|\mathbf{h}\| > \|(\mathbf{J} - K)\mathbf{h}\|$ , which is our desired contradiction.

As an example, if  $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  and  $\mathbf{x} \in \mathbb{R}^n$ , then the derivative of  $T$  at  $\mathbf{x}$  is  $T$ , as

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|T(\mathbf{x} + \mathbf{h}) - T\mathbf{x} - T\mathbf{h}\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|\mathbf{0}\|}{\|\mathbf{h}\|} = 0.$$

It is very intuitive to think of  $\mathbf{J}$  as an approximation of  $f$  at  $\mathbf{x}_0$  — namely, that there exists  $r(\mathbf{h})$  such that  $f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \mathbf{J}\mathbf{h} - r(\mathbf{h})$  and  $\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{r(\mathbf{h})}{\|\mathbf{h}\|} = 0$ . This strategy will be exhibited in the following proof:

## 2.2 Chain Rule

**Theorem.** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$ . If  $f$  is differentiable at  $\mathbf{x}_0$  and  $g$  is differentiable at  $f(\mathbf{x}_0)$  — and if  $\mathbf{x}_0$  and  $f(\mathbf{x}_0)$  are contained within open sets in the domains of  $f$  and  $g$  respectively — then  $g \circ f$  is differentiable at  $\mathbf{x}_0$ , and*

$$(g \circ f)' = g'(f(\mathbf{x}_0))f'(\mathbf{x}_0).$$

*Proof.* Let  $f'(\mathbf{x}_0) = \mathbf{J}$  and  $g'(f(\mathbf{x}_0)) = K$ . We have that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} = 0 = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|g(f(\mathbf{x}_0) + \mathbf{h}) - g(f(\mathbf{x}_0)) - K\mathbf{h}\|}{\|\mathbf{h}\|}.$$

Define the function  $\mathbf{k} = f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$ ; clearly,  $\lim_{\mathbf{h} \rightarrow \mathbf{0}} \mathbf{k} = \mathbf{0}$ . We have that

$$\begin{aligned}&g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - (K\mathbf{J})\mathbf{h} \\ &= g(f(\mathbf{x}_0) + \mathbf{k}) - g(f(\mathbf{x}_0)) - K(\mathbf{J}\mathbf{h} - f(\mathbf{x}_0 + \mathbf{h}) + f(\mathbf{x}_0) + \mathbf{k}) \\ &= g(f(\mathbf{x}_0) + \mathbf{k}) - g(f(\mathbf{x}_0)) - K\mathbf{k} + K(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h}).\end{aligned}$$



We now establish bounds for  $\|\mathbf{k}\|$ :

$$\|\mathbf{k}\| = \|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h} + \mathbf{J}\mathbf{h}\| \leq \|\mathbf{h}\| \left( \|\mathbf{J}\| + \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} \right)$$

Then using the composition rule for limits,

$$\begin{aligned} 0 &\leq \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) - (K\mathbf{J})\mathbf{h}\|}{\|\mathbf{h}\|} \\ &\leq \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|g(f(\mathbf{x}_0) + \mathbf{k}) - g(f(\mathbf{x}_0)) - K\mathbf{k}\|}{\|\mathbf{h}\|} + \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|K\| \|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} \\ &\leq \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\|g(f(\mathbf{x}_0) + \mathbf{k}) - g(f(\mathbf{x}_0)) - K\mathbf{k}\|}{\|\mathbf{k}\|} \lim_{\mathbf{h} \rightarrow \mathbf{0}} \left( \|\mathbf{J}\| + \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \mathbf{J}\mathbf{h}\|}{\|\mathbf{h}\|} \right) \\ &= (0)(\|\mathbf{J}\| + 0) = 0. \end{aligned}$$

so  $(g \circ f)'(\mathbf{x}_0) = K\mathbf{J} = g'(f(\mathbf{x}_0))f'(\mathbf{x}_0)$  as required. Yes, that last step of Rudin's is nonrigorous — define it as a piecewise expression equal to 0 whenever  $\mathbf{k} = \mathbf{0}$ , etc.

## 2.3 The Partial Derivative

Consider  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ , where  $U$  is an open subset of  $\mathbb{R}^n$ ; let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{e}_1, \dots, \mathbf{e}_m$  be the standard bases of  $\mathbb{R}^n$  and  $\mathbb{R}^m$ . The **components** of  $f$  are the real functions  $f_1, \dots, f_m$  defined by

$$f(\mathbf{x}) = f_1(\mathbf{x})\mathbf{e}_1 + \dots + f_m(\mathbf{x})\mathbf{e}_m,$$

or equivalently by  $f_i(\mathbf{x}) = f(\mathbf{x}) \cdot \mathbf{e}_i$  for each  $i \in \{1, \dots, m\}$ . Then for  $x \in U$ ,  $i \in \{1, \dots, m\}$ , and  $j \in \{1, \dots, n\}$ , we define the **partial derivative** of  $f_i$  with respect to  $x_j$  as

$$\frac{\partial f_i}{\partial x_j} = \lim_{t \rightarrow 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t}.$$

The fact this is a one-variable limit explains why the computation of partial derivatives aligns so closely with differentiation of univariate functions.

**Lemma.** *The entries of the total derivative are the partial derivatives: namely, if  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  (where  $U$  is open) and  $f$  is differentiable at  $x_0$ , then the partial derivatives exist and*

$$f'(\mathbf{x}_0)\mathbf{e}_j = \sum_{i=1}^m \left( \frac{\partial f_i}{\partial x_j} \right) (\mathbf{x}) \mathbf{e}_i$$

*Proof.* Let  $j$  be any integer in the set  $\{1, \dots, n\}$ . Since  $f$  is differentiable at  $\mathbf{x}$ ,

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t} - f'(\mathbf{x})\mathbf{e}_j = \lim_{t \rightarrow 0} \frac{\|f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0) - f'(\mathbf{x}_0)(t\mathbf{e}_j)\|}{\|t\mathbf{e}_j\|} = 0.$$

Therefore,

$$\begin{aligned}
f'(\mathbf{x})\mathbf{e}_j &= \lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t} \\
&= \lim_{t \rightarrow 0} \frac{\sum_{i=1}^m (f_1(\mathbf{x}_0 + t\mathbf{e}_j)\mathbf{e}_i) - \sum_{i=1}^m (f_i(\mathbf{x}_0)\mathbf{e}_i)}{t} \\
&= \sum_{i=1}^m \left( \lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t} \mathbf{e}_i \right) \\
&= \sum_{i=1}^m \left( \frac{\partial f_i}{\partial x_j} \right) (\mathbf{x}_0) \mathbf{e}_i,
\end{aligned}$$

as desired.

We denote  $D_{21}$  as the double partial derivative of  $f$  with respect to the first, then the second, variable.

## 2.4 Mixed Partial Derivatives

**Lemma.** Suppose  $f$  is defined in an open set  $U \subset \mathbb{R}^2$  and  $D_1$  and  $D_{21}$  exist at every point of  $U$ . Let  $Q \subset U$  be a closed rectangle with sides parallel to the coordinate axes with opposite vertices  $(a, b)$  and  $(a + h, b + k)$  for  $h, k \neq 0$ , and define

$$\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Then there is a point  $(x, y)$  in the interior of  $Q$  such that

$$\Delta(f, Q) = hk(D_{21}f)(x, y).$$

*Proof.* Define  $u(t) = f(t, b + k) - f(t, b)$ . Then by the Mean Value Theorem, there exists a  $x$  between  $a$  and  $a + h$  and  $y$  between  $b$  and  $b + K$  such that

$$\begin{aligned}
\Delta(f, Q) &= u(a + h) - u(a) \\
&= hu'(x) \\
&= h(D_1f(x, b + k) - D_1f(x, b)) \\
&= hkD_{21}f(x, y).
\end{aligned}$$

**Theorem 10.** Suppose  $f$  is defined in an open set  $U \in \mathbb{R}^2$ , that  $D_1$  and  $D_2$  exist at all points of  $U$ , and that  $D_{21}$  is continuous at some point  $(a, b) \in U$ . Then  $D_{12}$  exists at  $(a, b)$ , and

$$D_{12}f(a, b) = D_{21}f(a, b).$$

*Proof.*

## 2.5 Real-Valued Functions

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable real-valued function. Then  $f'$  is a 1-by- $n$  matrix:

$$f' = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

The transpose of this matrix is called the **gradient** of  $f$ ;

$$\nabla f = f'^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \frac{\partial f}{\partial x_1} \mathbf{e}_1 + \cdots + \frac{\partial f}{\partial x_n} \mathbf{e}_n,$$

where  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is the standard basis of  $\mathbb{R}^n$ . Observe that  $f' \mathbf{v} = \nabla f \cdot \mathbf{v}$  for all  $\mathbf{v} \in \mathbb{R}^n$  — so the gradient and the derivative of a real-valued function are dual concepts. We can define the derivative of  $f$  as a vector  $\nabla f$  such that

$$\lim_{\mathbf{h} \rightarrow 0} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \nabla f \cdot \mathbf{h}\|}{\|\mathbf{h}\|} = 0.$$

The **directional derivative** of  $f$  at  $\mathbf{x}$  along a unit vector  $\mathbf{v}$  is defined as

$$\nabla_{\mathbf{v}} f = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}$$

Observe that  $\nabla_{\mathbf{e}_i} f = \frac{\partial f}{\partial x_i} = \nabla f \cdot \mathbf{e}_i$  for all  $i \in \{1, \dots, n\}$ . This might lead us to conclude the following lemma:

**Lemma.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable at  $\mathbf{x}_0$ , then  $\nabla_{\mathbf{v}} f = \mathbf{v} \cdot (\nabla f)$  for all unit vectors  $\mathbf{v}$ .*

*Proof.* Observe that  $f' \mathbf{v} = \nabla f \cdot \mathbf{v}$ , so we may express the definition of the total derivative in terms of the gradient of  $f$ , and that  $\|t\mathbf{v}\| = |t|$ :

$$\begin{aligned} \nabla_{\mathbf{v}} f &= \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) - \nabla f \cdot (t\mathbf{v})}{t} + \lim_{t \rightarrow 0} \frac{\nabla f \cdot (t\mathbf{v})}{t} \\ &= 0 + \lim_{t \rightarrow 0} \nabla f \cdot \mathbf{v} \\ &= \nabla f \cdot \mathbf{v}, \end{aligned}$$

as required.

**Lemma.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable at  $\mathbf{x}_0$ , then the maximum of  $\nabla_{\mathbf{v}}f(\mathbf{x}_0)$  across all unit vectors  $\mathbf{v}$  occurs when  $\mathbf{v}$  points in the direction of  $\nabla f(\mathbf{x}_0)$ .*

*Proof.* If  $\mathbf{v}$  is a unit vector, then the Cauchy-Schwarz Inequality guarantees that

$$\nabla_{\mathbf{v}}f(\mathbf{x}_0) = \mathbf{v} \cdot \nabla f(\mathbf{x}_0) \leq \|\mathbf{v}\| \|\nabla f(\mathbf{x}_0)\| = \|\nabla f(\mathbf{x}_0)\| = \left( \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|} \right) \cdot \nabla f(\mathbf{x}_0)$$

so the maximum of  $\nabla_{\mathbf{v}}f(\mathbf{x}_0)$  occurs when  $\mathbf{v}$  is the normalization of the gradient vector and points in the direction of  $\nabla f(\mathbf{x}_0)$ .

More generally, we have that if  $\theta$  is the angle between the unit vector  $\mathbf{v}$  and  $\nabla f$ , then

$$\nabla_{\mathbf{v}}f = \mathbf{v} \cdot \nabla f = \|\mathbf{v}\| \|\nabla f\| \cos(\theta) = \|\nabla f\| \cos(\theta).$$

The directional derivative oscillates like a sine wave as  $\mathbf{v}$  walks around the unit hypersphere.

### 3 The Inverse Function Theorem

#### 3.1 The Contraction Principle

Let  $X$  be a metric space with metric  $d$ . If  $\varphi : X \rightarrow X$  and there exists a real  $c < 1$  such that

$$d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y)$$

for all  $x, y \in X$ , then  $\varphi$  is a **contraction** of  $X$  into  $X$ .

**Theorem.** *If  $X$  is a complete metric space and if  $\varphi$  is a contraction of  $X$  into  $X$ , then there exists a unique element  $x \in X$  such that  $\varphi(x) = x$*

*Proof.* Let  $c$  be a constant such that  $d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y)$  for all  $x, y \in X$ , and select from  $X$  some element  $x_0$ . We define the sequence  $x_0, x_1, \dots$  recursively by setting

$$x_{n+1} = \varphi(x_n)$$

for all  $n \in \mathbb{Z}_{\geq 0}$ . We have via induction that

$$d(x_{n+1}, x_n) = d(\varphi(x_n), \varphi(x_{n-1})) = c \cdot d(x_n, x_{n-1}) = \dots = c^n \cdot d(x_1, x_0).$$

We seek to invoke the completeness of  $X$ . Observe that for all  $N < n < m$ ,

$$\begin{aligned} d(x_m, x_n) &\leq d(x_{n+1}, x_n) + d(x_{n+2}, x_{n+1}) + \cdots + d(x_m, x_{m-1}) \\ &\leq (c^n + c^{n+1} + \cdots + c^{m-1})d(x_1, x_0) \\ &= c^n(1 + \cdots + c^{m-n-1})d(x_1, x_0) \\ &= \left( \frac{c^m - c^n}{c - 1} \right) d(x_1, x_0) \end{aligned}$$

As the right-hand side of this equation gets arbitrarily small (select  $N = \log_c(\epsilon)$  and let magic happen), we find that  $x_0, x_1, \dots$  is a Cauchy sequence. By completeness, it converges to some  $x \in X$ . Therefore,

$$x = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \varphi(x_n) = \varphi(f).$$

To prove that  $f$  is unique, note that if  $\varphi(x) = x$  and  $\varphi(y) = y$  for  $x, y \in X$ , then

$$d(x, y) = d(\varphi(x), \varphi(y)) \leq c \cdot d(x, y).$$

As  $c < 1$ , we must have that  $d(x, y) = 0$ , so  $x = y$ .

### 3.2 The Inverse Function Theorem

**Theorem.** Suppose that  $C$  is a  $C^1$  mapping of an open set  $E \subset \mathbb{R}^n$  to  $\mathbb{R}^n$ , that the matrix  $f'(\mathbf{a})$  is invertible for  $\mathbf{a} \in E$ , and define  $\mathbf{b} = f(\mathbf{a})$  — then there exist open sets  $U, V \subset \mathbb{R}^n$  such that  $\mathbf{a} \in U$ ,  $\mathbf{b} \in V$ , and  $f$  is a bijective mapping from  $U$  to  $V$  — and the inverse  $g : V \rightarrow U$  of  $f$  defined by  $g(f(\mathbf{x})) = \mathbf{x}$  for all  $\mathbf{x} \in U$  is  $C^1$ .

*Proof.* Let  $f'(\mathbf{a}) = \mathbf{J}$ , and let  $\lambda = \frac{1}{2\|\mathbf{J}^{-1}\|}$ , and let  $U$  be the open ball defined by all vectors  $\mathbf{x}$  such that

$$\|f'(\mathbf{x}) - \mathbf{J}\| < \lambda.$$

Further define  $V = f(U)$  (or more formally,  $\mathbf{y} \in \mathbb{R}^m \mid \exists \mathbf{x} \in U, \mathbf{x} = \mathbf{y}$ ). We must prove that  $f$  is invertible, that  $V$  is open, and that  $g$  is  $C^1$ .

**Invertability of  $f$ :** We now associate to each  $\mathbf{y} \in \mathbb{R}^n$  a function  $\varphi_{\mathbf{y}}$  defined by

$$\varphi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} + \mathbf{J}^{-1}(\mathbf{y} - f(\mathbf{x})).$$

Clearly,  $f(\mathbf{x}) = \mathbf{y}$  if and only if  $\varphi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$ . Since  $\varphi'_{\mathbf{y}}(\mathbf{x}) = I - \mathbf{J}^{-1}f'(\mathbf{x}) = \mathbf{J}^{-1}(\mathbf{J} - f'(\mathbf{x}))$  for all  $\mathbf{x}$ , we find that

$$\|\varphi'_{\mathbf{y}}(\mathbf{x})\| = \|\mathbf{J}^{-1}(\mathbf{J} - f'(\mathbf{x}))\| \leq \|\mathbf{J}^{-1}\| \|\mathbf{J} - f'(\mathbf{x})\| < \|\mathbf{J}^{-1}\| \lambda = \frac{1}{2}.$$

We use an above theorem to conclude that for all  $x_1, x_2 \in U$ ,

$$\|\varphi_{\mathbf{y}}(\mathbf{x}_1) - \varphi_{\mathbf{y}}(\mathbf{x}_2)\| < \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|.$$

We conclude that the Contraction Principle guarantees that  $\varphi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$  has exactly one solution — so  $f(\mathbf{x}) = \mathbf{y}$  has exactly one solution. We conclude that  $f$  is bijective (and thus invertible) over  $U$ .

**Openness of  $V$ :** For all  $\mathbf{y}_0 \in V$ , select  $\mathbf{x}_0$  such that  $f(\mathbf{x}_0) = \mathbf{y}_0$ , and let  $r$  be the radius of an open ball  $B_{\mathbf{x}_0}$  centered at  $\mathbf{x}_0$  contained within  $U$ . We claim that if  $\|\mathbf{y} - \mathbf{y}_0\| < \lambda r$ , then  $\mathbf{y} \in V$ .

We must construct  $\mathbf{x} \in U$  such that  $f(\mathbf{x}) = \mathbf{y}$ , which we do by proving that  $\varphi_{\mathbf{y}}$  is a contraction of  $B_{\mathbf{x}_0}$  into  $U$ . If  $\|\mathbf{y} - \mathbf{y}_0\| < \lambda r$ , observe that

$$\|\varphi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\| = \|\mathbf{J}^{-1}(\mathbf{y} - \mathbf{y}_0)\| < \|J^{-1}\|\lambda r < \frac{r}{2}.$$

Then if  $\mathbf{x} \in B$ ,  $\|\mathbf{x} - \mathbf{x}_0\| < r$ , so

$$\|\varphi_{\mathbf{y}}(\mathbf{x}) - \mathbf{x}_0\| \leq \|\varphi_{\mathbf{y}}(\mathbf{x}) - \varphi_{\mathbf{y}}(\mathbf{x}_0)\| + \|\varphi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\| < \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\| + \frac{r}{2} \leq r,$$

so  $\varphi_{\mathbf{y}}(\mathbf{x}) \in B_{\mathbf{x}_0}$ . We conclude that  $\varphi_{\mathbf{y}}$  is a contraction of the complete metric space  $B_{\mathbf{x}_0}$  into itself, so it must have some fixed point  $\mathbf{x} \in B_{\mathbf{x}_0}$  such that  $\varphi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$ . Then  $f(\mathbf{x}) = \mathbf{y}$ , so  $\mathbf{y} \in f(B_{\mathbf{x}_0}) \subset f(U) = V$ . Thus  $V$  is an open set.

**Smoothness of Inverse:** For all  $\mathbf{y} \in V$  and  $\mathbf{y} + \mathbf{k} \in V$ , there exists  $\mathbf{x}$  and  $\mathbf{x} + \mathbf{h} \in U$  such that  $\mathbf{y} = f(\mathbf{x})$  and  $\mathbf{y} + \mathbf{k} = f(\mathbf{x} + \mathbf{h})$ . Then

$$\begin{aligned} \|\mathbf{h} - \mathbf{J}^{-1}\mathbf{k}\| &= \|\mathbf{h} + J^{-1}(f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}))\| \\ &= \|\varphi_{\mathbf{y}}(\mathbf{x} + \mathbf{h}) - \varphi_{\mathbf{y}}(\mathbf{x})\| \\ &\leq \frac{1}{2}\|\mathbf{x} + \mathbf{h} - \mathbf{x}\| \\ &= \frac{1}{2}\|\mathbf{h}\|. \end{aligned}$$

Then  $\|\mathbf{J}^{-1}\mathbf{k}\| \geq \frac{1}{2}\|\mathbf{h}\|$ , so  $\mathbf{h} \leq 2\|J^{-1}\|\mathbf{k} = \frac{1}{\lambda}\mathbf{k}$ . We begin to now investigate the derivative: see that as

$$\|f'(\mathbf{a}) - \mathbf{J}\| \|J^{-1}\| < \lambda \|J^{-1}\| = \frac{1}{2} < 1$$

$f'(\mathbf{a})$  is invertible. Since

$$g(\mathbf{y} + \mathbf{k}) - g(\mathbf{y}) - f'(\mathbf{x})^{-1}\mathbf{k} = \mathbf{h} - f'(\mathbf{x})\mathbf{k} = -f'(\mathbf{x})^{-1}(f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - f'(\mathbf{x})\mathbf{h}),$$

we have that

$$\frac{\|g(\mathbf{y} + \mathbf{k}) - g(\mathbf{y}) - f'(\mathbf{x})^{-1}\mathbf{k}\|}{\|\mathbf{k}\|} \leq \frac{\|f'(\mathbf{x})^{-1}\|}{\lambda} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - f'(\mathbf{x})\mathbf{h}\|}{\mathbf{h}}.$$

As  $\mathbf{k} \rightarrow \mathbf{0}$ , we have that  $\mathbf{h} \rightarrow \mathbf{0}$  (this is nonrigorous; we'd need to define a piecewise function in the instance that  $\mathbf{k} = 0$ ). Then as the right-hand side of this inequality tends to 0, the left-hand side does by the Squeeze Theorem. Thus,

$$g'(\mathbf{y}) = f'(\mathbf{x})^{-1} = f(g(\mathbf{y}))^{-1}$$

Finally, note that as  $g$  is a continuous mapping of  $V$  onto  $U$ , that  $f'$  is a continuous mapping of  $U$  into  $\Omega$ , then  $(f')^{-1}$  is a continuous mapping of  $U$  into  $\Omega$ , so  $g'(\mathbf{y})$  is a continuous mapping of  $V$  into  $\Omega$ . This completes the proof of the most complex (and beautiful) theorem I've ever studied.

If we lessen the restriction that  $f$  need be  $C^1$ , the only part of the Inverse Function Theorem that fails is that  $g$  is  $C^1$ ; if  $f$  is merely differentiable, we may derive that  $g$  is differentiable too.

### 3.3 The Implicit Function Theorem