Rudin: Functions of Several Variables

James Pagan

October 2023

Contents

1 Limits of Linear Operators

1.1 Matrix Norm

The **norm** of a linear map $\mathbf{T} \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ is defined as follows:

$$\|\mathbf{T}\| = \sup\{\|\mathbf{Tz}\| \mid \mathbf{z} \in \mathbb{C}^n, \|\mathbf{z}\| \leq 1\}.$$

Notice that the matrix norm is nonnegative. The **critical vector** of $\mathbf{T}$ is the vector $\mathbf{z} \in \mathbb{C}^n$ such that $\|\mathbf{z}\| \leq 1$ and $\|\mathbf{Tz}\| = \|\mathbf{T}\|$; the critical vector always has norm 1. Naturally, $\|\mathbf{Tz}\| \leq \|\mathbf{T}\| \cdot \|\mathbf{z}\|$; since equality is attained, $\|\mathbf{Tz}\| \leq \lambda\mathbf{z}$ for $\lambda \in \mathbb{R}^n$ implies $\|\mathbf{T}\| \leq \lambda$.

Theorem 1. *If* $\mathbf{T}, \mathbf{S} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$*, then* $\|\mathbf{T} + \mathbf{S}\| \leq \|\mathbf{T}\| + \|\mathbf{S}\|$ *and* $\|\mathbf{TS}\| \leq \|\mathbf{T}\|\|\mathbf{S}\|$.

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

$$\|\mathbf{T} + \mathbf{S}\| = \|(\mathbf{T} + \mathbf{S})\mathbf{z}\| \leq \|\mathbf{Tz}\| + \|\mathbf{Sz}\| \leq \|\mathbf{T}\|\|\mathbf{z}\| + \|\mathbf{S}\|\|\mathbf{z}\| = \|\mathbf{T}\| + \|\mathbf{S}\|.$$

Now, let $\mathbf{w}$ be the critical vector of $\mathbf{TS}$. Then

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

This completes the proof. □

Theorem 2. *The matrix norm is a metric of* $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Proof. Suppose $\mathbf{T}, \mathbf{S} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. We must perform four rather routine calculations:

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

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

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

2. **Symmetry**: Notice that $(\mathbf{T} - \mathbf{S})\mathbf{z} = -(\mathbf{S} - \mathbf{T})\mathbf{z}$ for all $\mathbf{z} \in \mathbb{C}^n$. 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 $\mathbf{R} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,

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

We conclude that the matrix norm is a metric of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. □

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

1.2 Properties of Linear Maps

We will denote the canonical basis of $\mathbb{C}^n$ as $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Theorem 3. *If $\mathbf{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $\|\mathbf{T}\| < \infty$ and $\mathbf{T}$ is a uniformly continuous mapping of $\mathbb{R}^n$ into $\mathbb{R}^n$.*

Proof. Let $\mathbf{z} = (z_1, \ldots z_n)$ be the critical vector of $\mathbf{T}$. Then $|z_i| \leq 1$ for each i, so

$$\begin{aligned}
\|\mathbf{T}\| &= \|\mathbf{Tz}\| \\
&\leq |z_1|\|\mathbf{Te}_1\| + \cdots + |z_n|\|\mathbf{Te}_n\| \\
&\leq \|\mathbf{Te}_1\| + \cdots + \|\mathbf{Te}_n\| \\
&< \infty.
\end{aligned}$$

As per uniform continuity: realize that if $\epsilon > 0$, then $0 \leq \|\mathbf{z} - \mathbf{w}\| < \frac{\epsilon}{\|\mathbf{T}\|}$ implies

$$\|\mathbf{Tz} - \mathbf{Tw}\| \leq \|\mathbf{T}\| \cdot \|\mathbf{z} - \mathbf{w}\| < \|\mathbf{T}\| \left(\frac{\epsilon}{\|\mathbf{T}\|} \right) = \epsilon.$$

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

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

Theorem 4. *If $\mathbf{T} \in \Omega$, and $\mathbf{S} \in \mathcal{L}(\mathbb{C}^n)$, then*

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

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

Proof. Suppose $\mathbf{S} \notin \Omega$. Then there exists $\mathbf{z} \in \mathbb{C}^n$ of norm 1 such that $\mathbf{Sz} = \mathbf{0}$, so

$$\begin{aligned}
1 &= \|\mathbf{z}\| \\
&= \left\|\mathbf{T}^{-1}\mathbf{Tz}\right\| \\
&\leq \left\|\mathbf{T}^{-1}\right\| \cdot \|\mathbf{Tz}\| \\
&= \left\|\mathbf{T}^{-1}\right\| \cdot \|\mathbf{Sz} - \mathbf{Tz}\| \\
&\leq \left\|\mathbf{T}^{-1}\right\| \cdot \|\mathbf{S} - \mathbf{T}\| \cdot \|\mathbf{z}\| \\
&= \left\|\mathbf{T}^{-1}\right\| \cdot \|\mathbf{S} - \mathbf{T}\|.
\end{aligned}$$

Taking the contrapositive yields the desired result. □

Theorem 5. Ω *is an open subset of* $\mathcal{L}(\mathbb{C}^n)$*, and the bijection* $f : \mathbf{T} \to \mathbf{T}^{-1}$ *is continuous on* Ω*.*

Proof. Let $\mathbf{T} \in \Omega$. Since $\left\|\mathbf{T}^{-1}\right\|$ 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

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

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

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

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

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

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

1.3 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 $\mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n$; for any $\mathbf{T} \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$, define the components of $\mathbf{T}$ as t_{ij} for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. Then

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

Then via the Cauchy-Schwarz Inequality,

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

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 6. *$\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ equipped with the matrix norm is a complete metric space.*

Proof. Let the sequence $\mathbf{T}_1, \mathbf{T}_2, \ldots$ be a Cauchy sequence in $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$; declare that for all $\epsilon > 0$, there exists an integer N such that

$$N < i, j \implies \|\mathbf{T}_i - \mathbf{T}_j\| < \epsilon.$$

Observe that if $k \in \{1, \ldots, n\}$, then $N \leq i, j$ implies

$$\|\mathbf{T}_i \mathbf{e}_k - \mathbf{T}_j \mathbf{e}_k\| \leq \|\mathbf{e}_k\| \cdot \|\mathbf{T}_i - \mathbf{T}_j\| = \|\mathbf{T}_i - \mathbf{T}_j\| < \epsilon.$$

Thus $\mathbf{T}_1 \mathbf{e}_k, \mathbf{T}_2 \mathbf{e}_k, \cdots$ is a Cauchy Sequence in $\mathbb{C}^n$. We may thus define $\mathbf{T}$ as the unique matrix that maps $\mathbf{e}_k$ to the limit of $\mathbf{T}_1 \mathbf{e}_k, \mathbf{T}_2 \mathbf{e}_k, \ldots$ in $\mathbb{C}^m$.

For $i \in \mathbb{Z}_{>0}$, let ${}_i t_{jk}$ be the components of $\mathbf{T}_i$ and let t_{jk} be the components of $\mathbf{T}$. It is straightforward that $\lim\limits_{i \to \infty} {}_i t_{jk} = t_{jk}$; then there exist N_{jk} such that

$$N_{jk} < i \implies |{}_i t_{jk} - t_{jk}| < \frac{\epsilon}{\sqrt{mn}}.$$

Set $N = \max\{N_{11}, \ldots, N_{jk}\}$. Then $N < i$ implies that

$$\|\mathbf{T}_i - \mathbf{T}\| \leq \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} |{}_i t_{jk} - t_{jk}|^2} < \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} \frac{\epsilon^2}{mn}} = \epsilon.$$

We conclude that $\mathbf{T}_1, \mathbf{T}_2, \ldots$ converges to $\mathbf{T}$, so $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ is a complete metric space. $\square$

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 $X \to \mathbf{T}$ 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 $X \to \mathbf{T}$ is continuous. $\square$

2 Differentiation

2.1 The Derivative

Let $f : E \to \mathbb{R}^m$ 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

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

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 matricies $\mathbf{J} \neq K$ such that

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

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

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

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 $\|\mathbf{J} - K\|\|\mathbf{h}\| > \|(J - K)\mathbf{h}\|$, 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

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

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

2.2 Chain Rule

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$. 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

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

Define the function $\mathbf{k} = f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0)$; clearly, $\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{k} = 0$. 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 $\|\mathbf{k}\|$:

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

Then using the composition rule for limits,

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

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

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

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

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

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

Lemma. *The entries of the total derivative are the partial derivatives: namely, if $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ (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 $\{1, \ldots, n\}$. Since f is differentiable at $\mathbf{x}$,

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

Therefore,

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

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 $U \subset R^2$ 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 exes with opposite verticies (a, b) and $(a+h, b+k)$ for $h, k \neq 0$, and define*

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

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

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

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

$$\begin{aligned}
\triangle(f, Q) &= u(a+h) - u(a) \\
&= hu'(x) \\
&= h(D_1 f(x, b+k) - D_1 f(x, b)) \\
&= hkD_{21}f(x, y).
\end{aligned}$$

Theorem 7. *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 $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 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 ∇f such that

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

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

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

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

Lemma. *If $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\|t\mathbf{v}\| = |t|$:

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

as required.

Lemma. *If $f : \mathbb{R}^n \to \mathbb{R}$ 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

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

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 θ is the angle between the unit vector $\mathbf{v}$ and ∇f, then

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

The directional derivative oscilates 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 $\varphi : X \to X$ 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 φ is a **contraction** of X into X.

Theorem. *If X is a complete metric space and if φ 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 $x_0, x_1, \ldots$ recursively by setting

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

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

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

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 $x_0, x_1, \ldots$ is a Cauchy sequence. By completeness, it converges to some $x \in X$. Therefore,

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

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 $U, V \in \mathbb{R}^n$ such that $\mathbf{a} \in U$, $\mathbf{b} \in V$, and f is a bijective mapping from U to V — and the inverse $g : V \to U$ 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 $\lambda = \frac{1}{2\|J^{-1}\|}$, and let U be the open ball defined by all vectors $\mathbf{x}$ such that

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

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 $y \in \mathbb{R}^n$ 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 $\varphi'_{\mathbf{y}}(\mathbf{x}) = I - \mathbf{J}^{-1} f'(\mathbf{x}) = J^{-1}(J - f'(\mathbf{x})$ for all $\mathbf{x}$), we find that

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

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

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

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

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

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

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

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

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

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

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

$f'(\mathbf{a})$ 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

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

As $\mathbf{k} \to \mathbf{0}$, we have that $\mathbf{h} \to \mathbf{0}$ (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 Ω, then $(f')^{-1}$ is a continuous mapping of U into Ω, so $g'(\mathbf{y})$ is a continuous mapping of V into Ω. 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