Econ 897 Math Camp Part II

Differentiation in $\mathbb{R}$ and $\mathbb{R}^n$

Preliminary*

Author: David Zarruk Valencia

University of Pennsylvania

Summer 2016

Contents

---

*All comments and corrections are welcome. All errors are mine. Special thanks to Ju Hu, who contributed with many of the exercises.

1 Differentiation in $\mathbb{R}$

Recall the definition of continuity from the first part of the course:

Definition 1. *Let $f : [a, b] \to \mathbb{R}$, be a real-valued function, $p \in [a, b]$. We write $f(x) \to q$ as $x \to p$ or $\lim_{x \to p} f(x) = q$ if there exists $q \in \mathbb{R}$ with the following property: $\forall \epsilon > 0$ there exists $\delta > 0$ such that:*

$$|f(x) - q| < \epsilon, \quad \forall x \in (p - \delta, p + \delta) \cap [a, b], \quad x \neq p$$

The following theorem gives two different characterizations to the definition above:

Theorem 1. *Let $f : [a, b] \to \mathbb{R}$, be a real-valued function, $p \in [a, b]$. The following are equivalent:*

1. $\lim_{x \to p} f(x) = q$
2. $\lim_{n \to \infty} f(x_n) = q$, *for every sequence* $\{x_n\}_{n=0}^{\infty}, \quad x_n \neq p$ *such that* $\lim_{n \to \infty} x_n = p$
3. $\lim_{x \downarrow p} f(x) = \lim_{x \uparrow p} f(x) = q$

These equivalences can be proven with the material covered in the first part of the course and can be found on the reference books. They will be very useful when we want to show whether a function is differentiable or not. This will become clear soon.

Definition 2. *Let $f : [a, b] \to \mathbb{R}$, be a real-valued function. We say that f is differentiable at $x_0 \in [a, b]$ if the limit:*

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists and is finite. We denote it by:

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

and say that f' is the derivative of f at x_0. If f is differentiable at every $x \in [a, b]$, we say that f is differentiable.

Given Theorem 1 we can characterize the derivative of a function as:

Theorem 2. *Let $f : [a, b] \to \mathbb{R}$, be a real-valued function, $x_0 \in [a, b]$. The following are equivalent:*

1. *f is differentiable at x_0, with derivative $f'(x_0)$*
2. $\lim_{n\to\infty} \frac{f(x_n)-f(x_0)}{x_n-x_0} = f'(x_0)$, *for every sequence* $\{x_n\}_{n=0}^{\infty}, x_n \neq x_0$, *such that* $\lim_{n\to\infty} x_n = x_0$
3. $\lim_{x\downarrow x_0} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\uparrow x_0} \frac{f(x)-f(x_0)}{x-x_0} = f'(x_0)$

The following example computes the derivative of some common functions and shows some functions that are not differentiable.

Example 1.

1. *Let $c \in \mathbb{R}$ and $f(x) = c$. Then $f'(x) = 0, \forall x \in \mathbb{R}$*
2. *Let $n \geq 1$ and $f(x) = x^n$. Then $f'(x) = nx^{n-1}$*
3. *Let $f(x) = e^x$. Then $f'(x) = e^x$.*
4. *$f(x) = |x|$ is not differentiable at $x = 0$*
5. $f(x) = \begin{cases} x \cdot \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$ *is not differentiable at $x = 0$*

Proof.

1. $f(x) = c$:
$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x\to x_0} \frac{c - c}{x - x_0} = 0$$
2. $f(x) = x^n$:

$$\begin{aligned} f(x) - f(x_0) &= x^n - x_0^n \\ &= (x - x_0) \cdot (x^{n-1} + x^{n-2}x_0 + x^{n-3}x_0^2 + \ldots + x \cdot x_0^{n-2} + x_0^{n-1}) \\ &= (x - x_0) \sum_{k=0}^{n-1} x^k x_0^{n-k-1} \end{aligned}$$

Then,

$$\begin{aligned} f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} &= \lim_{x\to x_0} \sum_{k=0}^{n-1} x^k x_0^{n-k-1} \\ &= \sum_{k=0}^{n-1} x_0^k x_0^{n-k-1} \\ &= n x_0^{n-1} \end{aligned}$$

3. $f(x) = e^x$:

$$\begin{aligned} \lim_{x_0} \frac{f(x)-f(x_0)}{x-x_0} &= \lim_{x\to x_0} \frac{e^x - e^{x_0}}{x-x_0} \\ &= e^{x_0}\left(\lim_{x\to x_0} \frac{e^{x-x_0}-1}{x-x_0}\right) \\ &= e^{x_0}\left(\lim_{x\to 0} \frac{e^x-1}{x}\right) \end{aligned}$$

Recall the definition of e,

$$e := \lim_{y\to 0}(1+y)^{1/y}$$

Using the continuity of log at 1,

$$\lim_{y\to 0} \quad log(1+y)^{1/y} = \lim_{y\to 0} \quad \frac{1}{y} log(1+y) = 1$$

Using the continuity of $1/x$ at 1,

$$\lim_{y\to 0} \quad \frac{y}{log(1+y)} = 1$$

Defining $x = log(1+y)$,

$$\lim_{x\to 0} \frac{e^x-1}{x} = 1$$

Thus,

$$f'(x_0) = \lim_{x_0} \frac{f(x)-f(x_0)}{x-x_0} = e^{x_0}\left(\lim_{x\to 0}\frac{e^x-1}{x}\right) = e^{x_0}$$

4. $f(x) = |x|$:

We prove it using the characterizations given by Theorem 2:

$$\lim_{x\downarrow x_0} \frac{f(x)-f(0)}{x} = 1 \neq \lim_{x\uparrow x_0} \frac{f(x)-f(0)}{x} = -1$$

Given that the limits from above and from below are not equal to each other, the function is not differentiable at 0.

5. $f(x) = \begin{cases} x \cdot \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$:

$$\frac{f(x) - f(0)}{x} = \frac{x \sin(\frac{1}{x})}{x} = \sin\left(\frac{1}{x}\right)$$

Consider the sequences $x_n = \frac{1}{2n\pi}$ and $y_n = \frac{1}{2n\pi + \pi/2}$, $\quad n \in \mathbb{N}$.

$$\sin(x_n) = \sin(2n\pi) = 0, \quad n \in \mathbb{N}$$

$$\sin(y_n) = \sin(2n\pi + \pi/2) = 1, \quad n \in \mathbb{N}$$

By Theorem 2, f is not differentiable.

□

The following theorem states the relationship between continuity and differentiation of a function.

Theorem 3. *Let $f : [a, b] \to \mathbb{R}$, be a real-valued function, differentiable at $x_0 \in [a, b]$. Then f is continuous at x_0.*

Proof.

$$\begin{aligned} f(x) &= f(x) + f(x_0) - f(x_0) \\ &= f(x_0) + \left(\frac{f(x) - f(x_0)}{x - x_0}\right) \cdot (x - x_0) \end{aligned}$$

Taking limits when $x \to x_0$,

$$\lim_{x \to x_0} f(x) = f(x_0) + \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0}\right) \cdot (x - x_0)$$

Given that $\lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0}\right)$ exists and is finite and $\lim_{x \to x_0} x - x_0 = 0$,

$$\Rightarrow \quad \lim_{x \to x_0} f(x) = f(x_0)$$

□

Note that the converse of this theorem is not true. That is, if a function $f : [a, b] \to \mathbb{R}$ is continuous at some point $x_0 \in [a, b]$, it need not be differentiable at that point. For example, the function

$$f(x) = \begin{cases} x \cdot \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is continuous at 0, but is not differentiable at that point.

Theorem 4. *Suppose $f, g : [a, b] \to \mathbb{R}$, be real-valued functions, differentiable at $x_0 \in [a, b]$ and $k \in \mathbb{R}$. Then:*

1. $(kf)'(x_0) = kf'(x_0)$
2. $(f + g)'(x_0) = f'(x_0) + g'(x_0)$
3. $(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$
4. *if* $g(x_0) \neq 0$, *then* $\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - g'(x_0)f(x_0)}{g(x_0)^2}$

Proof. See Rudin (1976). □

Example 2. *Let $n \geq 1$ and $f(x) = x^n$. Then $f'(x) = nx^{n-1}$.*

Proof. By induction. For the base case, $n = 1$:

$$f(x) = x, \quad f'(x_0) = \lim_{x \to x_0} \frac{x - x_0}{x - x_0} = 1 = nx_0^{n-1}$$

Assume it holds for $n = k$:

$$f(x) = x^k, \quad f'(x_0) = kx_0^{k-1}$$

For $n = k + 1$:

$$f(x) = x^{k+1} = x \cdot x^k$$

Using the product rule in Theorem 4,

$$f'(x_0) = 1 \cdot x_0^k + x \cdot kx_0^{k-1} = x_0^k + kx_0^k = (k+1)x_0^k$$

□

Theorem 5 (Chain rule)**.** *Let I and J be two intervals in $\mathbb{R}$. Let $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$, $f(I) \subseteq J$. If f is differentiable at $x_0 \in I$ and g is differentiable at $f(x_0) \in J$, then:*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$$

Proof. See Rudin (1976). □

1.1 Mean Value Theorems

Definition 3. *Let $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$.*

1. *$x_0 \in A$ is a global maximum of f if $f(x_0) \geq f(x), \quad \forall x \in A$*
2. *$x_0 \in A$ is a global minimum of f if $f(x_0) \leq f(x), \quad \forall x \in A$*
3. *$x_0 \in A$ is a local maximum of f if $\exists \delta > 0$ such that $f(x_0) \geq f(x), \quad \forall x \in A \cap (x_0 - \delta, x_0 + \delta)$*
4. *$x_0 \in A$ is a local minimum of f if $\exists \delta > 0$ such that $f(x_0) \leq f(x), \quad \forall x \in A \cap (x_0 - \delta, x_0 + \delta)$*

Theorem 6. *Let $f : [a, b] \to \mathbb{R}$. If f has a local maximum (minimum) at $x_0 \in (a, b)$ and f is differentiable at x_0, then $f'(x_0) = 0$.*

Proof. Let f have a local maximum at $x_0 \in (a, b)$. Then, there exists a $\delta > 0$ such that:

- For all $x \in (x_0, x_0 + \delta)$:

$$\frac{f(x) - f(x_0)}{x - x_0} \leq 0 \quad \Rightarrow \lim_{x \downarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \leq 0$$

- For all $x \in (x_0 - \delta, x_0)$:

$$\frac{f(x) - f(x_0)}{x - x_0} \geq 0 \quad \Rightarrow \lim_{x \uparrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \geq 0$$

The limits exist, given that f is differentiable at x_0. By Theorem 2,

$$0 \leq f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \leq 0$$

□

Theorem 7 (Rolle's Theorem)**.** *$f : [a, b] \to \mathbb{R}$ continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $x \in (a, b)$ such that $f'(x) = 0$*

Proof. Define

$$x_1 = \arg\min_{x \in [a,b]} f(x), \quad m = \min_{x \in [a,b]} f(x)$$
$$x_2 = \arg\max_{x \in [a,b]} f(x), \quad M = \max_{x \in [a,b]} f(x)$$

- If $m = M$, f is constant and $f'(x) = 0, \forall x \in [a, b]$
- If $m < M$, at least one of x_1 or x_2 is different from both a and b, given that $f(x_1) < f(x_2)$ and $f(a) = f(b)$. Without loss of generality, assume $x_1 \in (a, b)$. By Rolle's Theorem (Theorem 7), $f'(x_1) = 0$.

□

Theorem 8 (Cauchy's Mean Value Theorem)**.** *Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous and differentiable on (a, b). There exists $x_0 \in (a, b)$ such that*

$$f'(x_0)(g(b) - g(a)) = g'(x_0)(f(b) - f(a))$$

Proof. Define $h(t) := f(t)(g(b) - g(a)) - g(t)(f(b) - f(a))$. h is continuous on $[a, b]$, differentiable on (a, b) and $h(a) = h(b)$. By Rolle's Theorem (Theorem 7), there exists an $x_0 \in (a, b)$ such that $h'(x_0) = 0$. This happens if, and only if,

$$f'(x_0)(g(b) - g(a)) = g'(x_0)(f(b) - f(a))$$

□

Theorem 9 (Mean Value Theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous and differentiable on (a, b). There exists $x_0 \in (a, b)$ such that*

$$f(b) - f(a) = f'(x_0)(b - a)$$

Proof. Set $g(x) = x$ in Cauchy's Mean Value Theorem (Theorem 8). □

Theorem 10. *Let $f : [a, b] \to \mathbb{R}$ is continuous, differentiable on (a, b) and*

$$\sup_{x \in (a,b)} |f'(x)| \leq M$$

Then,

$$|f(x) - f(x')| \leq M|x - x'|, \quad x, x' \in [a, b]$$

Proof. Let $x, x' \in [a, b], x < x'$. By Mean Value Theorem (Theorem 9) there exists $\zeta \in (x, x')$ such that

$$|f(x) - f(x')| = |f'(\zeta)| \cdot |x - x'| \leq M|x - x'|$$

□

Definition 4. *Let $f : I \to \mathbb{R}$. If for all $x_1, x_2 \in I$:*

1. *$x_1 \leq x_2 \Rightarrow f(x_1) \leq f(x_2)$, we say that f is monotonically increasing*
2. *$x_1 \leq x_2 \Rightarrow f(x_1) \geq f(x_2)$, we say that f is monotonically decreasing*
3. *$x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$, we say that f is strictly monotonically increasing*
4. *$x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$, we say that f is strictly monotonically decreasing*

The next theorem characterizes monotonic functions in terms of their derivatives:

Theorem 11. *Let $f : [a, b] \to \mathbb{R}$ continuous and differentiable on (a, b).*

1. *f is increasing on $(a, b) \iff f'(x) \geq 0, \forall x \in (a, b)$*
2. *f is decreasing on $(a, b) \iff f'(x) \leq 0, \forall x \in (a, b)$*
3. *f is strictly increasing on (a, b) if $f'(x) > 0, \forall x \in (a, b)$*
4. *f is strictly decreasing on (a, b) if $f'(x) < 0, \forall x \in (a, b)$*

Proof.

1. $\Rightarrow$: f is increasing $\Rightarrow$ for all $x < x'$, $\frac{f(x')-f(x)}{x'-x} \geq 0$. Taking limits:

   $$f'(x) = \lim_{x' \downarrow x} \frac{f(x') - f(x)}{x' - x} \geq 0$$

   $\Leftarrow$: $f'(x) \geq 0$ for all $x \in (a, b)$. Let $x_1 < x_2$. By the Mean Value Theorem, there exists $\zeta \in (x_1, x_2)$ such that:

   $$f(x_2) - f(x_1) = f'(\zeta)(x_2 - x_1) \geq 0$$

   Then, $f(x_2) \geq f(x_1)$.

2. Analogous to 1.

3. $f'(x) > 0$ for all $x \in (a, b)$. Let $x_1 < x_2$. By the Mean Value Theorem, there exists $\zeta \in (x_1, x_2)$ such that:

$$f(x_2) - f(x_1) = f'(\zeta)(x_2 - x_1) > 0$$

Then, $f(x_2) > f(x_1)$.

4. Analogous to 3.

□

Note that 3. and 4. go only in one direction: if the derivative is strictly positive (negative), the function is strictly increasing (decreasing). However, a function that is strictly increasing (decreasing) does not necessarily have strictly positive (negative) derivative at every point in the domain. An example of such a function is $f(x) = x^3$. In this case, f is strictly increasing, although $f'(0) = 0$.

Are derivatives continuous? Not necessarily. For example, the function:

$$f(x) = \begin{cases} x^2 \cdot \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is differentiable at every point. However, the derivative is not continuous at 0. Although we cannot claim that the derivative of a function is continuous, derivatives and continuous functions have something in common: they take on all the intermediate values.

Theorem 12 (Intermediate Value Theorem for Derivatives)**.** *Let $f : [a, b] \to \mathbb{R}$ continuous and differentiable on $[a, b]$. If $f'(a) < \lambda < f'(b)$, there exists $x \in (a, b)$ such that $f'(x) = \lambda$.*

Proof. Let λ such that $f'(a) < \lambda < f'(b)$. Define $g(t) := f(t) - \lambda t$. Then:

$$g'(t) = f'(t) - \lambda, \quad g'(a) < 0, \quad g'(b) > 0$$

This means that g is decreasing on a and increasing on b, so we can find $x_1, x_2 \in (a, b)$ such that $g(x_1) < g(a)$ and $g(x_2) < g(b)$. Thus, g attains a minimum at some x in the interior of $[a, b]$. By Theorem 6, $g'(x) = f'(x) - \lambda = 0$. Then:

$$f'(x) = \lambda$$

□

Theorem 13 (Inverse Function Theorem)**.** *Let $f : (a, b) \to (c, d)$ be surjective, continuous and differentiable on (a, b), and $f'(x) \neq 0, \forall x \in (a, b)$. Then f is a homeomorphism and its inverse f^{-1} is differentiable, with:*

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$$

Proof. If $f'(x) \neq 0, \forall x \in (a, b)$, by the Intermediate Value Theorem for Derivatives, $f'(x)$ is either positive for all $x \in (a, b)$, or negative. Assume, without loss of generality, that $f'(x) > 0, \forall x \in (a, b)$.

Let $a < x_1 < x_2 < b$. By the Mean Value Theorem, there exists $\zeta \in (x_1, x_2)$ such that:

$$f(x_2) - f(x_1) = f'(\zeta)(x_2 - x_1) > 0$$

Then, f is strictly monotonically increasing, so it is injective. Since, by assumption, it is also surjective, its inverse f^{-1} exists and is well defined. Moreover, since f is differentiable, it is continuous on (a, b).

Now, lets prove that a strictly monotonic and continuous function is a homeomorphism. Let $y_0 \in (c, d)$ and $\epsilon > 0$. Denote $x_0 = f^{-1}(y_0)$ and define $y^- = f(x_0 - \epsilon)$ and $y^+ = f(x_0 + \epsilon)$. Let $\delta = \min\{|y^+ - y_0|, |y^- - y_0|\}$.

Since f is monotonic, f^{-1} is also monotonic, so $f^{-1}(y_0 + \delta) \leq x_0 + \epsilon$, $f^{-1}(y_0 - \delta) \geq x_0 - \epsilon$ and $f^{-1}(y_0 - \delta, y_0 + \delta)$ is an interval. Moreover, f is continuous, so $f^{-1}(y_0 - \delta, y_0 + \delta)$ is an open set, which means that $f^{-1}(y_0 - \delta, y_0 + \delta) \subseteq (x_0 - \epsilon, x_0 + \epsilon)$, so f^{-1} is continuous and f is a homeomorphism.

Now, lets show that:

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$$

Let $x_0 = f^{-1}(y_0), \quad x = f^{-1}(y)$.

$$\begin{aligned}(f^{-1})'(y_0) &= \lim_{y\to y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} \\ &= \lim_{x\to x_0} \frac{x - x_0}{f(x) - f(x_0)} \\ &= \frac{1}{f'(x_0)} \\ &= \frac{1}{f'(f^{-1}(y_0))}\end{aligned}$$

The second equality is true because f^{-1} is continuous, which implies that $y \to y_0$ if and only if $x \to x_0$. □

Example 3. *Let* $y = sin(x), \quad x \in (-\pi/2, \pi/2)$. *Find* $(f^{-1})'(y)$.

Proof. $f^{-1}(y) = arcsin(y)$. Then, by the Inverse Function Theorem:

$$(f^{-1})'(y) = \frac{1}{cos(arcsin(x))} = \frac{1}{\sqrt{1 - sin^2(arcsin(y))}} = \frac{1}{\sqrt{1 - y^2}}$$

□

1.2 L'Hospital's Rule

Theorem 14 (L'Hospital's Rule)**.** *Suppose f and g are differentiable on (a,b), $g'(x) \neq 0, \forall x \in (a,b)$, where $-\infty \leq a \leq b \leq \infty$. Suppose:*

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = A, \quad -\infty \leq A \leq \infty$$

If either:

1. $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$
2. $\lim_{x\to a} g(x) = \infty$

Then, $\lim_{x\to a} \frac{f(x)}{g(x)} = A$

Proof. Without loss of generality, assume $-\infty \leq A < \infty$. Let $A < r < q$. Since:

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = A$$

There exists c such that:

$$\frac{f'(x)}{g'(x)} < r, \quad \forall x \in (a,c)$$

By Cauchy's Mean Value Theorem, let $a < x < y < c$. Then:

$$\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f'(t)}{g'(t)} < r, \quad t \in (x,y) \subseteq (a,c)$$

If 1. holds, then:

$$\lim_{x\to a} \frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f(y)}{g(y)} \leq r < q$$

That is, for every $q > A$ there exists c such that $\frac{f(y)}{g(y)} < q$ for every $y \in (a,c)$.

If 2. holds, there exists $c_1 > a$ such that $g(x) > g(y)$ and $g(x) > 0$ for all $x \in (a, c_1)$.

$$\Rightarrow \quad \frac{f(x)-f(y)}{g(x)-g(y)} \cdot \frac{g(x)-g(y)}{g(x)} < r \cdot \frac{g(x)-g(y)}{g(x)}$$

$$\Rightarrow \quad \frac{f(x)}{g(x)} < r + \frac{f(y)}{g(x)} - r \cdot \frac{g(y)}{g(x)}$$

Then, there exists c_2 such that $\frac{f(x)}{g(x)} < q, \quad \forall x \in (a, c_2)$. □

1.3 Higher Order Derivatives and Taylor's Theorem

Definition 5 (Higher Order Derivatives)**.** *Let $f : [a, b] \to \mathbb{R}$ be differentiable.*

- *If f' is continuous, we say f is continuously differentiable and denote it as $f \in C^1$.*
- *If f' is differentiable, we say that f is twice-differentiable, and denote the second derivative as f''. If, in addition, f'' is continuous, we say that f is twice-continuously differentiable and denote it $f \in C^2$.*
- *If $f^{(n)}$ is differentiable, we say that f is $(n+1)$th-differentiable, and denote the $(n+1)$th derivative as $f^{(n+1)}$. If, in addition, $f^{(n+1)}$ is continuous, we say that f is $(n+1)$th-continuously differentiable and denote it $f \in C^{(n+1)}$.*

Theorem 15 (Taylor's Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be r-th order differentiable. Define:*

$$P(h) := f(x) + f'(x)h + \ldots + \frac{f^{(r)}(x)h^r}{r!}$$

Then:

1. $\lim_{h\to 0} \frac{f(x+h)-P(h)}{x^r} = 0$
2. *$P(h)$ is the only polynomial of degree lower than or equal to r with this approximation property*
3. *If, in addition, f is $(r+1)$-th order differentiable, there exists $\zeta \in (x, x+h)$ such that:*

$$f(x+h) = P(h) + \frac{f^{(r+1)}(\zeta)h^{r+1}}{(r+1)!}$$

Proof. 1. Define the residual function $R(h) := f(x+h) - P(h)$. We want to show that:

$$\lim_{h\to 0} \frac{R(h)}{h^r} = 0$$

Note that $R(0) = f(x) - P(0) = 0$. By the Mean Value Theorem, there exists $\theta_1 \in (0, h)$ such that:

$$R(h) = R(h) - R(0) = R'(\theta_1)(h - 0) = R'(\theta_1)h$$

Similarly, $R'(0) = f'(x) - P'(0) = 0$, so there exists $\theta_2 \in (0, \theta_1)$ such that:

$$R(h) = (R'(\theta_1) - R'(0))h = R''(\theta_2)(\theta_1 - 0)h = R''(\theta_2)\theta_1 h$$

Continuing in the same fashion:

$$R(h) = \ldots = R^{(r-1)}(\theta_{r-1})\theta_{r-2}\theta_{r-3}\ldots\theta_2\theta_1 h, \quad 0 < \theta_{r-1} < \ldots < \theta_1 < h$$

That is, $\{\theta_n\}_{n=0}^{r-1}$ is decreasing, so:

$$\begin{aligned}
\left|\frac{R(h)}{h^r}\right| &= \left|\frac{R^{(r-1)}(\theta_{r-1})\theta_{r-2}\theta_{r-3}\ldots\theta_2\theta_1 h}{h^r}\right| \\
&\leq \left|\frac{R^{(r-1)}(\theta_{r-1})h^{r-1}}{h^r}\right| \\
&= \left|\frac{R^{(r-1)}(\theta_{r-1})}{h}\right| \\
&\leq \left|\frac{R^{(r-1)}(\theta_{r-1})}{\theta_{r-1}}\right| \xrightarrow{h\to 0} 0
\end{aligned}$$

Then, $\lim_{h\to 0}\left|\frac{R(h)}{h^r}\right| = 0$.

2. Let:

$$P(h) = a_0 + a_1 h + \ldots + a_r h^r$$

$$Q(h) = b_0 + b_1 h + \ldots + b_r h^r$$

Suppose $P \neq Q$ are two polynomials such that:

$$\lim_{h\to 0}\frac{f(x+h) - P(h)}{h^r} = 0$$

$$\lim_{h\to 0}\frac{f(x+h) - Q(h)}{h^r} = 0$$

Then:

$$\frac{f(x+h) - Q(h)}{h^r} = \frac{f(x+h) - P(h)}{h^r} + \frac{P(h) - Q(h)}{h^r} = 0$$

Which means that $\lim_{h\to 0}\frac{P(h)-Q(h)}{h^r} = 0$.

There exists $0 \leq k \leq r$ such that $a_k \neq b_k$. Let k_0 the highest such k.

- If $k_0 = r$, then $\lim_{h\to 0}\frac{P(h)-Q(h)}{h^r} = a_r - b_r \neq 0$
- If $k_0 < r$, then $\lim_{h\to 0}\frac{P(h)-Q(h)}{h^r} = \pm\infty$

3. $R(h) = f(x+h) - P(h)$. Define $g(h) := h^{r+1}$.

$$\begin{aligned}
\frac{R(h)}{g(h)} &= \frac{R(h) - R(0)}{g(h) - g(0)} \\
&= \frac{R'(\theta_1) - R'(0)}{g'(\theta_1) - g'(0)} \\
&= \dots \\
&= \frac{R^{(r+1)}(\theta_{r+1})}{(r+1)!} \\
&= \frac{f^{(r+1)}(\theta_{r+1})h^{r+1}}{(r+1)!}
\end{aligned}$$

□

1.4 Exercises

1. Let $\alpha \in \mathbb{R}$ and

$$f(x) = \begin{cases} x^\alpha \cdot \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

State whether f is differentiable or not. Does it depend on the value of α?

2. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be twice differentiable. Suppose there exists $\epsilon > 0$ such that $f''(x) > \epsilon$ for all $x \in \mathbb{R}$. Show that $f'(x) = 0$ for some $x \in \mathbb{R}$.

3. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be twice continuously differentiable. Assume there is a $c \in (a, b)$ such that $f'(c) = 0$ and $f''(c) < 0$. Show that f has a local maximum at c.

4. Suppose that $f : (a, b) \longrightarrow \mathbb{R}$ is differentiable and f' is bounded. If $\{x_n\}$ is a sequence on (a, b) and $x_n \longrightarrow a$, then $f(x_n)$ converges.

5. Show that $e^x > 1 + x$ for all $x > 0$.

6. Show that if $\alpha > 1$, then $(1 + x)^\alpha > 1 + \alpha x$ for all $x > 0$. Similarly, it $\alpha < 1$, $(1 + x)^\alpha < 1 + \alpha x$ for all $x > 0$.

7. Let $f : (a, b) \to \mathbb{R}$ be differentiable and f' increasing. Show that f' is continuous.

8. Give an example of a function f that is differentiable, but whose derivative f' is not continuous.

9. Let $f : [a, b] \to \mathbb{R}$ continuous and differentiable. Assume $f(a) < 0$, $f(b) < 0$, $f(c) > 0$, where $a < c < b$. Prove that there exists $\zeta \in (a, b)$ such that $f(\zeta) + f'(\zeta) = 0$.

10. Show that $e^x = ax^2 + bx + c$ has at most 3 real roots.

11. Let $f(a, b) :\to \mathbb{R}$. Assume that f is differentiable at $x_0 \in (a, b)$. Show that

$$\lim_{h \to 0} \frac{f(x_0 + h^2) - f(x_0)}{h}$$

exists.

12. Let:

$$f(x) = \begin{cases} x^2 & \text{if } x \notin \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{Q} \end{cases}$$

Is f differentiable at $x = 0$?

13. Let $f : \mathbb{R} \to \mathbb{R}$ be third-order differentiable. Assume that $\sup_{x\in\mathbb{R}} |f(x)| \leq M_1$, $\sup_{x\in\mathbb{R}} |f'''(x)| \leq M_2$. Then, f' and f'' are bounded.

14. Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions.

    (a) Assume f is differentiable at x_0 but g is not differentiable at x_0. Prove $f(x)+g(x)$ is not differentiable at x_0.

    (b) Assume both f and g are not differentiable at x_0. Can $f(x)+g(x)$ be differentiable at x_0?

15. Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions. Let $y_0 = g(x_0)$ for some $x_0 \in \mathbb{R}$. Consider the following cases:

    (a) g is differentiable at x_0 and f is not differentiable at y_0;

    (b) g is not differentiable at x_0 and f is differentiable at y_0;

    (c) g is not differentiable at x_0 and f is not differentiable at y_0.

    For each case, find examples of f and g such that $f \circ g$ is differentiable at x_0.

16. Assume f is differentiable at some x_0. Calculate the following two limits.

    (a) $\lim_{h\to 0} \frac{f(x_0-h)-f(x_0)}{h}$;

    (b) $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0-h)}{h}$.

17. (Exercise 1 on page 186, Pugh)

18. (Exercise 5 on page 186, Pugh)

19. (Exercise 11 on page 186, Pugh) Assume that $f : (-1, 1) \to \mathbb{R}$ and $f'(0)$ exists. If $\alpha_n, \beta_n \to 0$ as $n \to \infty$, define the different quotient

$$D_n = \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n}.$$

    (a) Prove that $\lim_{n\to\infty} D_n = f'(0)$ under each of the following conditions

        i. $\alpha_n < 0 < \beta_n$.

        ii. $0 < \alpha_n < \beta_n$ and $\frac{\beta_n}{\beta_n-\alpha_n} \leq M$.

        iii. $f'(x)$ exists and is continuous for all $x \in (-1, 1)$.

(b) Set $f(x) = x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$. Observe that f is differentiable everywhere in $(-1, 1)$ and $f'(0) = 0$. Find α_n and β_n that tend to 0 in such a way that D_n converges to a limit unequal to $f'(0)$.

20. (Exercise 13 on page 187, Pugh) Assume that $f : \mathbb{R} \to \mathbb{R}$ is differentiable.

    (a) If there is an $L < 1$ such that for each $x \in \mathbb{R}$, $f'(x) < L$, prove that there exists a unique point x such that $f(x) = x$.

    (b) Show by example that (a) fails if $L = 1$.

21. Suppose $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable. Assume $f(0) > 0$, $f'(0) < 0$ and $f''(x) < 0$ for all $x \in \mathbb{R}$. Prove there exists $\xi \in \left(0, -\frac{f(0)}{f'(0)}\right)$ such that $f(\xi) = 0$.

22. Assume $f : [a, b] \to \mathbb{R}$ is twice differentiable and $f'(a) = f'(b) = 0$. Prove there exists $\xi \in (a, b)$ such that

$$\left|f''(\xi)\right| \geq \frac{4}{(b-a)^2}\left|f(b) - f(a)\right|.$$

(Hint: expand $f\left(\frac{a+b}{2}\right)$ at a and b respectively)

23. (Exercise 10 on page 187, Pugh) Let $f : (a, b) \to \mathbb{R}$ be given.

    (a) If $f''(x)$ exists, prove that

$$\lim_{h \to 0} \frac{f(x-h) - 2f(x) + f(x+h)}{h^2} = f''(x).$$

    (b) Find an example that this limit can exist even when $f''(x)$ fails to exist.

24. Let $f : [a, b] \to \mathbb{R}$ be twice differentiable. Assume $\sup_{x \in [a,b]} |f''(x)| \leq M$ for some constant M. Assume also f achieves its global maximum at some point x^* in (a, b). Prove

$$\left|f'(a)\right| + \left|f'(b)\right| \leq M(b-a).$$

25. Assume f function is continuous on $[0, \infty)$ and differentiable on $(0, \infty)$. Suppose $f(0) = 0$ and f' is increasing on $(0, \infty)$. Prove

$$g(x) = \frac{f(x)}{x}$$

is increasing on $(0, \infty)$.

2 Differentiation in $\mathbf{R}^n$

2.1 Preamble

The following section assumes knowledge of linear algebra. Particularly, the following theorems will be used henceforth.

Definition 6. *Let V and W be vector spaces. The mapping $T : V \to W$ is linear if $\forall v, v' \in V, \alpha, \beta \in \mathbb{R}$:*

$$T(\alpha v + \beta v') = \alpha T(v) + \beta T(v')$$

Definition 7. *Let V and W be vector spaces. We say that V and W are isomorphic if there exists a linear mapping $T : V \to W$ that is bijective. T is called an isomorphism.*

Theorem 16. *Let V and W be vector spaces and $\dim(V) < \infty, \dim(W) < \infty$. Then, V and W are isomorphic if, and only if, $\dim(V) = \dim(W)$.*

The importance of this theorem relies on the fact that if we want to study *any* finite dimensional vector space of dimension n, it suffices to study $\mathbb{R}^n$.

2.2 $\mathcal{L}(V, W)$ as a normed space

Recall that a norm on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$ such that, for every $v, w \in V$ and $\lambda \in \mathbb{R}$:

1. $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$
2. $\|\lambda v\| = |\lambda| \cdot \|v\|$
3. $\|v + w\| \leq \|v\| + \|w\|$

We are now going to endow the vector space $\mathcal{L}(V, W)$ of all linear maps from V to W with a norm.

Definition 8. *Let V and W be two normed vector spaces with $\|\cdot\|_V$ and $\|\cdot\|_W$ the respective norms. Consider the map $T : V \to W$. That is, $T \in \mathcal{L}(V, W)$. Define the operator norm $\|\cdot\|$ on $\mathcal{L}(V, W)$ as:*

$$\|T\| = \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V}$$

Note that for $v \in V$, the term $\frac{\|Tv\|_W}{\|v\|_V}$ is the "stretch" of the vector v after T is applied to it. Therefore, the operator norm is the supremum of the "stretches" of the the vectors in V under the operator T.

Example 4. *1.* $T : \mathbb{R} \to \mathbb{R}$, $T(x) = \alpha x, \alpha \in \mathbb{R}$.

$$\frac{\|Tx\|}{\|x\|} = \frac{|\alpha x|}{|x|} = \frac{|\alpha| \cdot |x|}{|x|} = |\alpha|, \quad \forall x \in \mathbb{R}$$

$$\Rightarrow \|T\| = |\alpha|$$

2. $T : \mathbb{R} \to \mathbb{R}^n$, $T(x) = vx, v \in \mathbb{R}^n$

$$\frac{\|Tx\|}{\|x\|} = \frac{\|vx\|}{|x|} = \frac{\|v\| \cdot |x|}{|x|} = \|v\|, \quad \forall x \in \mathbb{R}$$

$$\Rightarrow \|T\| = \|v\|$$

The following theorem endows the vector space $\mathcal{L}(V, W)$ with the operator norm $\|\cdot\|$ defined above.

Theorem 17. *$\|\cdot\|$ is a norm on $\mathcal{L}(V, W)$.*

Proof. Left as an exercise. □

The next theorem gives another characterizations of the operator norm. In particular, it states that to compute the norm of an operator $T : V \to W$, it suffices to find the maximum stretch of the operator over the elements on the unit sphere, instead of the stretch of every single vector in V. The intuition behind the proof is simple: any linear operator has exactly the same "stretch" over all the multiples of a vector v. That is:

$$\frac{\|Tv\|}{\|v\|} = \frac{\|T(\lambda v)\|}{\|\lambda v\|}, \quad \forall \lambda \in \mathbb{R}$$

This means that computing the supremum over all the elements of V is equivalent to finding the supremum over all the elements in the unit sphere in V.

Theorem 18.

$$\|T\| = \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V} = \sup_{\substack{v \neq 0 \\ \|v\| \leq 1}} \frac{\|Tv\|_W}{\|v\|_V} = \sup_{\|v\|=1} \|Tv\|_W$$

Proof. Define:

$$a := \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V}, \quad b := \sup_{\substack{v \neq 0 \\ \|v\| \leq 1}} \frac{\|Tv\|_W}{\|v\|_V}, \quad c := \sup_{\|v\|=1} \|Tv\|_W$$

We are going to show that $a \geq b \geq c \geq a$.

- $a \geq b$: $\{v : \|v\| \leq 1, v \neq 0\} \subseteq \{v : v \neq 0\}$. Therefore, the supremum taken over a bigger set must be greater than or equal.

- $b \geq c$: $\{v : \|v\| = 1, v \neq 0\} \subseteq \{v : \|v\| \leq 1, v \neq 0\}$. Therefore, the supremum taken over a bigger set must be greater than or equal.

- $C \geq a$:

  Assume $c < a$. This means:

$$\sup_{\|v\|=1} \|Tv\|_W < \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V}$$

  $\Rightarrow \exists v_0 \in V$ such that:

$$\sup_{\|v\|=1} \|Tv\|_W < \frac{\|Tv_0\|_W}{\|v_0\|_V}$$

$$\frac{\|Tv_0\|_W}{\|v_0\|_V} = \left(\frac{1}{\|v_0\|_V}\right) \cdot \|Tv_0\|_W = \left\|\left(\frac{1}{\|v_0\|_V}\right) \cdot Tv_0\right\|_W = \left\|T\left(\frac{v_0}{\|v_0\|_V}\right)\right\|_W$$

  Denote $z := \frac{v_0}{\|v_0\|_V}$. Clearly, $\|z\| = 1$, which means that:

$$\frac{\|Tv_0\|_W}{\|v_0\|_V} = \|Tz\|_W > \sup_{\|v\|=1} \|Tv\|_W, \quad z \in V, \|z\| = 1$$

  which is a contradiction. Thus, $a = b = c$.

□

The following theorems are properties of the operator norm $\|\cdot\|$ that will be used later when studying derivatives in $\mathbb{R}^n$.

Theorem 19 (Cauchy-Schwartz Inequality)**.** $\|Tv\|_W \leq \|T\| \cdot \|v\|_V\,, \quad \forall v \in V$

Proof. If $v = 0$, then $\|v\| = 0$ and $\|Tv\| = 0$.
If $v \neq 0$:

$$\frac{\|Tv\|_W}{\|v\|_V} \leq \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V} = \|T\| \quad \Rightarrow \quad \|Tv\|_W \leq \|T\| \cdot \|v\|_V\,, \quad \forall v \in V$$

□

Theorem 20. *If* $\|Tv\|_W \leq \lambda \|v\|_V\,, \quad \forall v \in V$ *and* $\lambda > 0$*, then:*

$$\|T\| \leq \lambda$$

Proof.

$$\frac{\|Tv\|_W}{\|v\|_V} \leq \lambda, \quad \forall v \in V$$

Taking the supremum over all $v \in V$, the inequality remains. □

Theorem 21. *If* $T : V \to W$ *and* $S : W \to U$ *are linear maps between normed spaces, then:*

$$\|S \circ T\| \leq \|S\| \, \|T\|$$

Proof. Using Theorem 19:

$$\frac{\|(S \circ T)v\|_W}{\|v\|_V} = \frac{\|S(T(v))\|_W}{\|v\|_V} \leq \frac{\|S\| \cdot \|Tv\|_W}{\|v\|_V} \leq \frac{\|S\| \, \|T\| \, \|v\|_V}{\|v\|_V} = \|S\| \, \|T\|$$

□

Now that we have endowed $\mathcal{L}(V, W)$ with a norm, we can define a metric in the natural way:

$$d(T, S) = \|T - S\|\,, \quad \forall T, S \in \mathcal{L}(V, W)$$

With a metric, we can talk about open and closed sets in $\mathcal{L}(V, W)$ and apply all the topological properties of metric spaces derived in the first part of the math camp.

The following theorem will prove to be very useful later on. It gives another characterization for continuous operators T.

Theorem 22. *Let $T \in \mathcal{L}(V, W)$ be a linear mapping. The following are equivalent:*

1. $\|T\| < \infty$
2. *T is uniformly continuous*
3. *T is continuous*
4. *T is continuous at the origin*

Proof. 1. $\Rightarrow$ 2.: Assume $\|T\| < \infty$.

$$\left\|Tv - Tv'\right\|_W = \left\|T(v - v')\right\|_W \leq \|T\| \left\|v - v'\right\|_V$$

Let $\epsilon > 0$ and $\delta = \epsilon / \|T\|$. For all $v, v' \in V$ such that $\|v - v'\|_V < \delta = \epsilon / \|T\|$:

$$\left\|Tv - Tv'\right\|_W \leq \|T\| \left\|v - v'\right\|_V < \|T\| \cdot \frac{\epsilon}{\|T\|} = \epsilon$$

Thus, T is uniformly continuous.

2. $\Rightarrow$ 3.: Immediate.

3. $\Rightarrow$ 4.: Immediate.

4. $\Rightarrow$ 1.: Assume T is continuous at the origin and take $\epsilon = 1$. $\exists \delta > 0$ such that $\forall v \in V$:

$$\|v\|_V < \delta \quad \Rightarrow \quad \|T(v) - T(0)\|_W = \|Tv\|_W < \epsilon = 1$$

Let $v \in V$ and set $u := \lambda v$, where $\lambda = \frac{\delta}{2\|v\|_V}$.:

$$\|u\|_V = \|\lambda v\|_V = \frac{\delta}{2} < \delta \quad \Rightarrow \quad \|Tu\|_W < \epsilon = 1$$

$$\frac{\|Tv\|_W}{\|v\|_V} = \frac{\|Tu\|_W}{\|u\|_V} < \frac{1}{\|u\|_V} = \frac{2}{\delta}$$

Taking supremum over all $v \in V$:

$$\|T\| \leq \frac{2}{\delta} < \infty$$

$\square$

The idea behind the proof of the last implication is the following. The stretch of any vector v under T is the same that the stretch of a multiple of that vector under T. Thus, for any v, we take a vector u, which is a multiple of v that has a sufficiently small norm. We know that T is continuous at the origin, so the image of u under T will be bounded, proving that its stretch is finite.

Theorem 23. *Let $T : \mathbb{R}^n \to W$ be a linear mapping to W, a normed vector space. Then T is continuous. If, in addition, T is an isomorphism, the T is a homeomorphism.*

Proof. First, we will show that any linear mapping from $\mathbb{R}^n$ to any normed vector space W is continuous. By Theorem 22, It suffices to show that $\|T\| < \infty$.

$$\|v\| = \sqrt{v_1^2 + \ldots + v_n^2} \geq |v_i|, \quad i \in \{1, \ldots, n\}$$

Let $M = \max\{|Te_1|, \ldots, |Te_n|\}$ where e_i are the unit vectors in $\mathbb{R}^n$.

$$\begin{aligned} \|Tv\|_W &= \|T(v_1 e_1 + \ldots + v_n e_n)\|_W \\ &\leq |v_1| \|Te_1\| + \ldots + |v_n| \|Te_n\| \\ &\leq (|v_1| + \ldots + |v_n|) \cdot M \\ &\leq nM \|v\| \end{aligned}$$

$$\Rightarrow \frac{\|Tv\|_W}{\|v\|} \leq nM, \quad \Rightarrow \quad \|T\| = \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V} \leq nM < \infty$$

This proves that T is continuous. Now, assume T is an isomorphism. We want to show that T^{-1} is continuous. That is, $\left\|T^{-1}\right\| < \infty$. Define the unit sphere in $\mathbb{R}^n$:

$$S^{n-1} = \{v \in \mathbb{R}^n : \|v\| = 1\}$$

S^{n-1} is compact in $\mathbb{R}^n$. Since T is continuous, $T(S^{n-1})$ is compact in W. Given that T is an isomorphism:

$$T(v) = 0 \iff v = 0$$

So, since $0 \notin S^{n-1} \Rightarrow 0 \notin T(S^{n-1})$ and $\exists c > 0$ such that $\|w\| > c, \forall w \in T(S^{n-1})$.

It can be easily shown that the preimage of elements w inside the $c-$sphere in W lie strictly inside S^{n-1}. That is:

$$\forall w \in W, w \notin T(S^{n-1}), \|w\|_W < c \quad \Rightarrow \quad \left\|T^{-1}w\right\| < 1$$

Now, lets to the same as in the proof of Theorem 22. Take any $w \in W$ and define $u = \lambda w$, where $\lambda = \frac{c}{2\|w\|_W}$.

$$\|u\|_W < c \quad \Rightarrow \quad \left\|T^{-1}u\right\| = \frac{\left\|T^{-1}w\right\|}{\|w\|_W} \cdot \frac{c}{2} < 1$$

$$\Rightarrow \quad \frac{\left\|T^{-1}w\right\|}{\|w\|_W} < \frac{2}{c} < \infty \quad \Rightarrow \quad \left\|T^{-1}\right\| < \infty$$

□

Theorem 24. *Let $T : V \to W$ be a linear mapping. If $\dim(V) = n < \infty$ then T is continuous and if T is an isomorphism, T is an homeomorphism.*

Proof. $\dim(V) = n$ so V and $\mathbb{R}^n$ are isomorphic. Let $H : \mathbb{R}^n \to V$ be an isomorphism. By Theorem 23, H is a homeomorphism, so H^{-1} is continuous. Moreover, $T \circ H : \mathbb{R}^n \to W$ is also continuous. Therefore, $T = (T \circ H) \circ H^{-1}$ is continuous.

If T is an isomorphism, since H is isomorphism, $T \circ H$ is an isomorphism and, by Theorem 23, a homeoporphism. That is, $(T \circ H)^{-1} = (H^{-1} \circ T^{-1})$ is continuous. Then, $T^{-1} = H \circ (H^{-1} \circ T^{-1})$ is continuous.

□

3 Derivatives in $\mathbb{R}^n$

Recall that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}$ if the following limit exists and is finite:

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

We say that the derivative of f at x is:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

This is equivalent to saying that f is differentiable at x, with derivative $f'(x)$, if there exists a function $r : \mathbb{R} \to \mathbb{R}$ such that:

$$f(x+h) - f(x) = f'(x) \cdot h + r(h)$$

And the remainder r is "sublinear":

$$\lim_{h \to 0} \frac{r(h)}{h} = 0$$

Note that, for a given x, the term $f'(x)h$ is linear in h, so we can interpret the derivative $f'(x)$ not as a number, but as a linear operator in $\mathbb{R}$, that maps h to $f'(x)h$. This is a natural way to extend the concept of derivative to $\mathbb{R}^n$:

Definition 9. *Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$. The function f is differentiable at $p \in U$, if there exists a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that:*

$$f(p+v) - f(p) = T(v) + R(v)$$

and the remainder function R is sublinear:

$$\lim_{v \to 0} \frac{\|R(v)\|}{\|v\|} = 0$$

We say that the derivative (also called total derivative of Frechet derivative) is $(Df)_p = T$.

This is equivalent to saying that $f : U \to \mathbb{R}^m$ is differentiable at $p \in U$ if there exists a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{v \to 0} \frac{\|f(x+v) - f(x) - T(v)\|}{\|v\|} = 0$$

Theorem 25. *If f is differentiable at $p \in U$, then the derivative is uniquely determined by:*

$$(Df)_p(u) = \lim_{t\to 0} \frac{f(p+tu)-f(p)}{t}$$

Proof. Let T be a linear map satisfying $f(p+v)-f(p) = T(v)+R(v)$ and $\lim_{v\to 0} \frac{\|R(v)\|}{\|v\|} = 0$.

$$\begin{aligned}
\lim_{t\to 0} \frac{f(p+tu)-f(p)}{t} &= \lim_{t\to 0} \frac{T(tu)}{t} + \frac{R(tu)}{t} \\
&= \lim_{t\to 0} \frac{tT(u)}{t} + \frac{R(tu)}{t} \\
&= T(u) + \lim_{t\to 0} \frac{R(tu)}{t\,\|u\|} \cdot \|u\|
\end{aligned}$$

Given that $\|u\|$ is finite and R is sublinear, the second term vanishes, so:

$$\lim_{t\to 0} \frac{f(p+tu)-f(p)}{t} = T(u)$$

Since limits are unique, if there are two such transformations T and T', they must be equal to each other: $T = T'$. □

Now, we state some of the theorems we saw in the univariate case, extended for the multivariate case.

Theorem 26. *Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$. Suppose f is differentiable at p. Then f is continuous at p.*

Proof. $(Df)_p : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, from $\mathbb{R}^n$ to a normed vector space $\mathbb{R}^m$. By Theorem 23, $(Df)_p$ is continuous. This is equivalent to $\|(Df)_p\| < \infty$.

$$\begin{aligned}
\lim_{v\to 0} \|f(p+v)-f(p)\| &= \lim_{v\to 0} \|(Df)_p(v) + R(v)\| \\
&\leq \lim_{v\to 0} \|(Df)_p\| \cdot \|v\| + \|R(v)\| \\
&= 0
\end{aligned}$$

given that $\|(Df)_p\| < \infty$, $\lim_{v\to 0} \|v\| = 0$ and $\lim_{v\to 0} \|R(v)\| = 0$. □

Theorem 27. *Let $f, g : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$ be differentiable at $p \in U$, $\alpha \in \mathbb{R}$. Then:*

1. $(D(f+\alpha g))_p = (Df)_p + \alpha(Dg)_p$
2. *If $f(p) = c$, for all $p \in U$, then $(Df)_p = 0$*

3. *If* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is a linear mapping,* $f(v) = Av, A \in \mathbb{R}^m \times \mathbb{R}^n$*, then* $(Df)_p = A$ *for all* $p \in U$

Proof. Left as an exercise. □

Theorem 28 (Chain Rule)**.** *Let* $U \subseteq \mathbb{R}^n$ *and* $W \subseteq \mathbb{R}^m$ *be open sets. Let* $f : U \to \mathbb{R}^m$ *be differentiable at* $p \in U$ *and* $f(U) \subseteq W$*. Let* $g : W \to \mathbb{R}^l$ *be differentiable at* $f(p) \in W$*. Define* $h = g \circ f$*. Then* h *is differentiable at* $p \in U$ *and* $(Dh)_p = (Dg)_{f(p)} \cdot (Df)_p$

Proof.

$$\begin{aligned} f(p+v) - f(p) &= (Df)_p(v) + R(v) \\ g(f(p)+u) - g(f(p)) &= (Dg)_{f(p)}(u) + S(u) \\ g(f(p+v)) &= g(f(p) + (Df)_p(v) + R(v)) \\ &= g(f(p)) + (Dg)_{f(p)}((Df)_p(v) + R(v)) + S((Df)_p(v) + R(v)) \end{aligned}$$

Therefore,

$$\begin{aligned} g(f(p+v)) - g(f(p)) &= (Dg)_{f(p)}((Df)_p(v) + R(v)) + S((Df)_p(v) + R(v)) \\ &= (Dg)_{f(p)}(Df)_p(v) + (Dg)_{f(p)}R(v) + S((Df)_p(v) + R(v)) \end{aligned}$$

It now suffices to show that the last two terms are sublinear:

1. $(Dg)_{f(p)}R(v)$:

$$\lim_{v \to 0} \frac{\left\|(Dg)_{f(p)}R(v)\right\|}{\|v\|} \leq \lim_{v \to 0} \left\|(Dg)_{f(p)}\right\| \cdot \frac{\|R(v)\|}{\|v\|} = 0$$

as the first term is finite anr R is sublinear.

2. $S((Df)_p(v) + R(v))$:

$$\lim_{v \to 0} \frac{\|S((Df)_p(v) + R(v))\|}{\|v\|} = \lim_{v \to 0} \frac{\|S((Df)_p(v) + R(v))\|}{\|(Df)_p(v) + R(v)\|} \cdot \frac{\|(Df)_p(v) + R(v)\|}{\|v\|}$$

The limit when $v \to 0$ of the last term is finite:

$$\frac{\|(Df)_p(v) + R(v)\|}{\|v\|} \leq \frac{\|(Df)_p(v)\|}{\|v\|} + \frac{R(v)}{\|v\|} \leq \frac{\|(Df)_p\| \, \|v\|}{\|v\|} + \frac{R(v)}{\|v\|} = \|(Df)_p\| + \frac{R(v)}{\|v\|}$$

□

Theorem 29. *Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$. Then, f is differentiable at $p \in U$ if and only if each of its components f_i is differentiable at p. Furthermore, the derivative of the i-th component is the i-th component of the derivative.*

Proof. $\Rightarrow$: Let f be differentiable and define the projection on the i-th dimension as:

$$\pi_i : \mathbb{R}^n \to \mathbb{R}, \quad \pi_i(w_1, \ldots, w_i, \ldots, w_n) = w_i$$

Clearly, π_i is linear, so it is differentiable. Then, $f_i = \pi_1 \circ f$ is differentiable and:

$$(Df_i)_p = (D\pi_i)_{f(p)}(Df)_p$$

Moreover, since the projection can be represented by the $1 \times n$ vector that has 1 in the i-th component and 0 elsewhere:

$$A = (0, \ldots, 1, \ldots, 0)$$

We know that $(D\pi_i)_{f(p)}$ is represented by the matrix A. So:

$$(Df_i)_p = \pi_i \circ (Df)_p$$

$\Leftarrow$: Suppose each f_i is differentiable, with derivative $(Df_i)_p$. Construct:

$$A = \begin{bmatrix} (Df_1)_p \\ \vdots \\ (Df_m)_p \end{bmatrix}$$

$$\Rightarrow \quad f(p+h) - f(p) - (Df)_p \cdot h = \begin{bmatrix} f_1(p+h) - f_1(p) - (Df_1)_p \cdot h \\ \vdots \\ f_m(p+h) - f_m(p) - (Df_m)_p \cdot h \end{bmatrix}$$

Taking limits, this converges if and only if each each component converges. Therefore, $(Df)_p$ is the derivative of f. □

This theorem is important, because it shows that what makes calculus in $\mathbb{R}^n$ different from calculus in $\mathbb{R}$ is the multidimensionality of the domain, and not of the range.

Theorem 30 (Mean Value Theorem)**.** *Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$. Assume f is differentiable on U and the segment $[p, q]$ is contained in U. Then:*

$$|f(q) - f(p)| \leq M\,|q - p|\,, \quad M = \sup_{x \in U}\{\|(Df)_x\|\}$$

Proof. Assume the segment $[p.q]$ is contained in U. The segment can be parameterized as:

$$p + t(q - p), \quad t \in [0.1]$$

Define:

$$g : [0, 1] \to \mathbb{R}, \quad g(t) := (f(p) - f(q))^t \cdot f(p + t(q - p))$$

$$\Rightarrow \quad g'(t) = (f(p) - f(q))^t (Df)_{p+t(q-p)}(q - p)$$

By the Mean Value Theorem in $\mathbb{R}$, there exists $\zeta \in (0, 1)$ such that:

$$g(1) - g(0) = g'(\zeta) = (f(p) - f(q))^t (Df)_{p+\zeta(q-p)}(q - p)$$

$$g(1) - g(0) = (f(p) - f(q))^t \cdot (f(p) - f(q)) = -\|f(p) - f(q)\|^2$$

$$\Rightarrow \quad \|f(p) - f(q)\|^2 = (f(p) - f(q))^t (Df)_{p+\zeta(q-p)}(p - q)$$

By the Cauchy-Schwartz Inequality:

$$\|f(p) - f(q)\| \leq \left\|(Df)_{p+\zeta(q-p)}\right\| \cdot \|p - q\| \leq M\,\|p - q\|$$

□

Corollary 1. *Assume U is connected. Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$ be differentiable and $(Df)_x = 0$. Then f is constant.*

Proof. Let $x \in U$. Define $P(x) := \{y \in U | f(x) = f(y)\}$. Lets show that $P(x)$ is open:
Let $y \in P(x)$. Since U is open, there exists an ϵ-neighborhood of y, $O_y \subseteq U$, which is open. Let $z \in O_y$. The segment $[y, z] \subseteq O_y$. Then, $|f(y) - f(z)| \leq M\,|y - z| = 0$. This implies that $f(x) = f(y) = f(z)$ for every $z \in O_y$. Then $z \in P(x)$, which implies $O_y \subseteq P(x)$, so $P(x)$ is open.

Now we show $P(x) = U, \forall x \in U$. Assume $P(x) \neq U$. That is, assume there exists $x \in U, P(x) \neq U$. $P(x)$ and $\cup_{y \notin P(x)} P(y)$ are both open, disjoint and $U = P(x) \cup \left(\cup_{y \notin P(x)} P(y)\right)$. This implies that U is disconnected, which is a contradiction. Therefore, $P(x) = U$. □

3.1 Partial Derivatives

Definition 10. *Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$. Define the ij-th partial derivative of f at p as:*

$$\frac{\partial f_i(p)}{\partial x_j} = \lim_{t \to 0} \frac{f_i(p + te_j) - f_i(p)}{t}$$

Theorem 31. *Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$ be differentiable. Then, the partial derivatives exist and are the entries of the matrix that represents the total derivative.*

Proof. Recall that the total derivative $(Df)_p$ is a linear map. This means that there exists a matrix of size $m \times n$ that represents $(Df)_p$. Let A be the matrix that represents the derivative $(Df)_p$. Then:

$$(Df)_p(e_j) = Ae_j = \lim_{t \to 0} \frac{f(p + te_j) - f(p)}{t} = \begin{bmatrix} \frac{\partial f_1(p)}{\partial x_j} \\ \vdots \\ \frac{\partial f_m(p)}{\partial x_j} \end{bmatrix}$$

Then:

$$A = \begin{bmatrix} \frac{\partial f_1(p)}{\partial x_1} & \cdots & \frac{\partial f_1(p)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(p)}{\partial x_1} & \cdots & \frac{\partial f_m(p)}{\partial x_n} \end{bmatrix}$$

□

Note that Theorem 31 states that if the derivative exists, then the partials also exist. A natural question is whether the converse is true. If the partial derivatives exist, is the function f differentiable? The following example shows that this is not the case.

Example 5. *Let:*

$$f(x) = \begin{cases} 0 & \textit{if } x, y = 0 \\ \frac{xy}{x^2+y^2} & \textit{otherwise} \end{cases}$$

f is not continuous at $(x, y) = (0, 0)$. To see this, take:

$$(x_n, y_n) = \left(\frac{1}{n}, \frac{1}{\sqrt{n}}\right) \xrightarrow{n \to \infty} (0, 0)$$

$$f(x_n, y_n) = \frac{1}{2}, \quad \forall n \geq 1$$

But $f(0, 0) = 0$, so f is not continuous. However, the partials exists. Note, however, that the partials are not continuous.

In the above example, we saw that the existence of the partials is not sufficient for the function to be differentiable. In particular, the partial derivatives of the function in the example existed, but were not continuous. The following theorem states a sufficient condition for f to be differentiable.

Theorem 32. *Let $f : U \to \mathbb{R}^m, U \subseteq \mathbb{R}^n$. If the partial derivatives of f exist and are continuous then f is differentiable.*

Proof. Assume the partials exist and are continuous. Without loss of generality, assume that $m = 1$ (Theorem 29). Let $h \in \mathbb{R}^n$.

$$\begin{aligned}
f(x+h) - f(x) &= f(x_1 + h_1, \ldots, x_n + h_h) - f(x_1, \ldots, x_n) \\
&= f(x_1 + h_1, \ldots, x_n + h_h) - f(x_1, x_2 + h_2, \ldots, x_n + h_n) \\
&+ f(x_1, x_2 + h_2, \ldots, x_n + h_h) - f(x_1, x_2, x_3 + h_3 \ldots, x_n + h_n) \\
&+ f(x_1, x_2, x_3 + h_3, \ldots, x_n + h_h) - f(x_1, x_2, x_3, x_4 + h_4, \ldots, x_n + h_n) \\
& \ldots \\
&+ f(x_1, x_2, \ldots, x_{n-1}, x_n + h_h) - f(x_1, x_2, \ldots, x_n)
\end{aligned}$$

We are "moving" component by component on each line. Using the Mean Value Theorem:

$$\begin{aligned}
&= \frac{\partial f}{\partial x_1}(\theta_1, x_2 + h_2, \ldots, x_n + h_n) h_1 \\
&+ \frac{\partial f}{\partial x_2}(x_1, \theta_2, x_3 + h_3, \ldots, x_n + h_n) h_2 \\
&+ \ldots \\
&+ \frac{\partial f}{\partial x_n}(x_1, \ldots, x_{n-1}, \theta_n) h_n
\end{aligned}$$

where $\theta_1 \in (x_1, x_1 + h_1), \ldots, \theta_n \in (x_n, x_n + h_n)$. Then:

$$\begin{aligned}
f(x+h) &- f(x) - A \cdot h \\
&= \left(\frac{\partial f}{\partial x_1}(\theta_1, x_2 + h_2, \ldots, x_n + h_n) - \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x_1, \ldots, x_{n-1}, \theta_n) - \frac{\partial f}{\partial x_n}(x) \right) \cdot h \\
&= z(h) \cdot h
\end{aligned}$$

By Cauchy-Schwartz Inequality:

$$\frac{\|f(x+h) - f(x) - A \cdot h\|}{\|h\|} \le \frac{\|z(h)\|}{\|h\|} \|h\| = \|z(h)\| \xrightarrow{h \to 0} 0$$

where the last inequality follows because the partials are continuous. Therefore, f is differentiable. $\square$

3.2 Higher Order Derivatives

Recall that for $f : U \to \mathbb{R}^m$ differentiable:

$$(Df) : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

$$x \mapsto (Df)_x$$

where, for all $x \in U$, $(Df)_x : U \to \mathbb{R}^m$. We define the second derivative analogously.

Definition 11. *(Df) is differentiable at $x \in \mathbb{R}^n$ if there exists a linear mapping*

$$T : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$$

such that:

$$(Df)_{x+v} - (Df)_x = T(v) + R(v), \quad \lim_{\|v\| \to 0} \frac{\|R(v)\|}{\|v\|} = 0$$

We denote the derivative as $(D^2 f) := T$ and call it the second derivative of f.

Note that $(D^2 f) : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, so when we evaluate the derivative at a point $p \in \mathbb{R}^n$, $(D^2 f)_p$ is a linear transformation from $\mathbb{R}^n$ to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. This means that for $v \in \mathbb{R}^n$, $(D^2 f)_p(v)$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$(D^2 f)_p(v) : \mathbb{R}^n \to \mathbb{R}^m$$

This finally means that:

$$(D^2 f)_p(v)(w) \in \mathbb{R}^m, \quad v \in \mathbb{R}^n, w \in \mathbb{R}^n$$

We can also interpret $(D^2 f)_p$ as a function:

$$(D^2 f)_p : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$$

$$(v, w) \mapsto (D^2 f)_p(v)(w) \in \mathbb{R}^m$$

Definition 12. *Let V, W, Z be vector spaces. A map $T : V \times W \to Z$ is bilinear if for every $v \in V$ and every $w \in W$, the maps $T(v, .) : W \to Z$ and $T(., w) : V \to Z$ are linear.*

Note that the second derivative $(D^2f)_p$ is a bilinear map.

Equivalently, (Df) is differentiable at $p \in \mathbb{R}^n$ if there exists a linear map $T : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ such that:

$$\lim_{\|v\|\to 0} \frac{\|(Df)_{p+v} - (Df)_p - Tv\|}{\|v\|} = 0$$

Theorem 33. *Let $T : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear map. There exists a unique matrix representation $A_{m\times n}$ of T, such that:*

$$g(u, v) = u^t Av, \quad u \in \mathbb{R}^m, v \in \mathbb{R}^n$$

Proof. See Lang (2010). □

That is, if the function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, there exists a matrix representation for the second derivative.

Theorem 34. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$. If $(D^2f)_p$ exists, then $(D^2f_k)_p$ exists, the second partial derivatives at p exist and*

$$(D^2f_k)_p(e_i)(e_j) = \frac{\partial^2 f_k(p)}{\partial x_i \partial x_j}$$

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable, so $(D^2f_k)_p$ exists. Let the mapping $S : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mu_{m\times n}$ be the isomorphism that assigns to a linear transformation its matrix representation:

$$S(T) = A, \quad T(v) = Av, \quad v \in \mathbb{R}^n$$

$(Df)_x$ is a linear transformation and is differentiable, so $S \circ (Df)_x$ is differentiable. Note that:

$$M_x := S \circ (Df)_x = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Note that the map $S \circ (Df) : \mathbb{R}^n \to \mu_{m\times n}$, and $\mu_{m\times n}$ is isomorphic to $\mathbb{R}^{m\times n}$. By Theorem 29, the map $S \circ (Df)_x$ is differentiable at $x = p$ if and only if all of the entries of M_x are differentiable. Then, its partial derivatives exist and are the derivatives of the entries of M_x:

$$\frac{\partial \left(\frac{\partial f_i}{\partial x_j}\right)}{\partial x_k} = \frac{\partial^2 f_i}{\partial x_j \partial x_k}$$

The second partial derivatives are $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$. □

Now, we can define the hessian. Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that $(D^2 f)_p$ exists. Then its representation matrix exists, we denote it the *hessian* matrix, and is given by:

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

Theorem 35. *If $(D^2 f)_p$ exists, it is symmetric:*

$$(D^2 f)_p(v)(w) = (D^2 f)_p(w)(v)$$

Proof. Without loss of generality, assume $m = 1$ (as symmetry concerns only the arguments of f, not its values). Let $f : \mathbb{R}^n \to \mathbb{R}$. Fix $v, w \in \mathbb{R}^n$. Let $t \in [0, 1]$ and define $g : [0, 1] \to \mathbb{R}$, where:

$$g(s) = f(p + tv + stw) - f(p + stw)$$

$$\Rightarrow \quad g(0) = f(p + tv) - f(p), \quad g(1) = f(p + tv + tw) - f(p + tw)$$

Note that $g(1) - g(0)$ is a symmetric function of v and w, so we can interchange them and get exactly the same result:

$$g(1) - g(0) = f(p + tv + tw) - f(p + tv) - f(p + tw) + f(p)$$

By the Mean Value Theorem, $g(1) - g(0) = g'(\theta), \theta \in (0, 1)$. By definition of the derivative:

$$(Df)_{p+tv+\theta tw} - (Df)_p = (D^2 f)_p(tv + \theta tw) + R(tv + \theta tw)$$

$$(Df)_{p+\theta tw} - (Df)_p = (D^2 f)_p(\theta tw) + S(\theta tw)$$

$$\Rightarrow \quad (Df)_{p+tv+\theta tw} - (Df)_{p+\theta tw} = (D^2f)_p(tv) + R(tv + \theta tw) - S(\theta tw)$$

$$g'(\theta) = (D^2f)_p(tv)(tw) + R(tv + \theta tw)(tw) - S(\theta tw)(tw)$$

$$\begin{aligned}
\frac{g(1) - g(0)}{t^2} &= \frac{g'(\theta)}{t^2} \\
&= \frac{(D^2f)_p(tv)(tw)}{t^2} + \frac{R(tv + \theta tw)(tw)}{t^2} + \frac{S(\theta tw)(tw)}{t^2} \\
&= (D^2f)_p(v)(w) + \frac{R(tv + \theta tw)(w)}{t} + \frac{S(\theta tw)(w)}{t} \\
&\xrightarrow{t \to 0} (D^2f)_p(v)(w)
\end{aligned}$$

The result is exactly the same if we interchange v and w:

$$\Rightarrow \quad (D^2f)_p(v)(w) = (D^2f)_p(w)(v)$$

$\square$

Corollary 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that f is twice differentiable. Then, there exists a symmetric matrix representation (hessian) for $(D^2f)_x$.*

Theorem 36 (Inverse Function Theorem)**.** *Let $f : U \to \mathbb{R}^n$, where $U \subseteq \mathbb{R}^n$ is open and f is continuously differentiable. Assume $(Df)_{x_0}$ is invertible for $x_0 \in U$ and $f(x_0) = y_0$. Then:*

1. *There exist neighborhoods $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^n$, $x_0 \in V$, $y_0 \in W$ such that $f : V \to W$ is a bijection.*

2. *If $g : W \to V$ is the inverse of f defined on W, where $g(f(x)) = x$, $x \in V$, then g is continuously differentiable and $(Dg)_{y_0} = (Df)^{-1}_{x_0}$*

Proof. See Rudin (1976). $\square$

Theorem 37 (Implicit Function Theorem)**.** *Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^{n+m}$, be a continuously differentiable mapping such that $f(x_0, y_0) = z_0$, $(x_0, y_0) \in U$, $x_0 \in \mathbb{R}^n$ and $y_0 \in \mathbb{R}^m$. Let:*

$$B := \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{bmatrix} \quad A := \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

If B is invertible, then there exists $V \subseteq \mathbb{R}^{n+m}$ and $W \in \mathbb{R}^n$ open sets, $(x_0, y_0) \in V$, $x_0 \in W$, such that for all $x \in W$ there exists a unique y, $(x, y) \in V$, such that $f(x, y) = z_0$. If y is defined as an implicit function of x, $y = g(x)$, then $g : W \to \mathbb{R}^m$ is continuously differentiable, $g(x_0) = y_0$, $f(x, g(x)) = z_0$ for all $x \in W$ and $(Dg)_{x_0} = -B^{-1}A$.

3.3 Exercises

1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xy$. Let $p = (p_1, p_2) \in \mathbb{R}^2$. Show that $(Df)_p = (p_2, p_1)$ is the derivative of f at p.

2. Prove Theorem 27.

3. Let $g : \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) = xy$ and $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f(x, y) = (x + y, y)$. Find the derivative of $g \circ f$.

   (Hint: Use the chain rule and the linearity of f).

References

Lang, S. (2010). Linear algebra. *Springer*.

Rudin, W. (1976). Principles of mathematical analysis. *McGraw-Hill International Editions*.