Notes for Vector Calculus

Zhao Wenchuan

November 21, 2021

Contents

Chapter 1.

Differentiation

§1.1 Differentiable Mapping

Definition 1.1.1 (Differentiable Mappings)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^n$.

f is said to be *differentiable* at a point $\mathbf{p} \in \mathbb{R}^m$ iff for any $\mathbf{t} \in \mathbb{R}^m \setminus \{\mathbf{0}_{\mathbb{R}^m}\}$, there exists a linear mapping $\phi : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$f(\mathbf{p} + \mathbf{t}) = f(\mathbf{p}) + \phi(\mathbf{t}) + o(\mathbf{t}) \quad \text{as } \mathbf{t} \to \mathbf{0}_{\mathbb{R}^m}.$$

Equivalently, that is

$$\lim_{\mathbf{t} \to \mathbf{0}_{\mathbb{R}^m}} \frac{f(\mathbf{p} + \mathbf{t}) - f(\mathbf{p}) - \phi(\mathbf{t})}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \mathbf{0}_{\mathbb{R}^n}.$$

Note 1.1.1. The equivalence of the assertions in Definition 1.1.1 can be proved as following.

$$\lim_{\mathbf{t} \to \mathbf{0}_{\mathbb{R}^m}} \frac{f(\mathbf{p} + \mathbf{t}) - f(\mathbf{p}) - \phi(\mathbf{t})}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \lim_{\mathbf{t} \to \mathbf{0}_{\mathbb{R}^m}} \frac{o(\mathbf{t})}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \mathbf{0}_{\mathbb{R}^m}.$$

Theorem 1.1.1. In Definition 1.1.1, ϕ is unique.

Proof. Suppose there exists another linear mapping $\lambda : \mathbb{R}^m \to \mathbb{R}^n$ such that

■

Theorem 1.1.2. With the condition in Definition 1.1.1, if f is differentiable at $\mathbf{p}$, then f is continuous at $\mathbf{p}$.

Proof. As f is differentiable at $\mathbf{p}$, there exists an $\alpha : \mathbb{R}^m \to \mathbb{R}^n$ with

$$\lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} \alpha(\mathbf{t}) = \alpha(\mathbf{0}_{\mathbb{R}^m}) = \mathbf{0}_{\mathbb{R}^m},$$

such that

$$\frac{f(\mathbf{p}+\mathbf{t}) - f(\mathbf{p}) - \phi(\mathbf{t})}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \alpha(\mathbf{t}).$$

By rearranging the equation, we observe

$$\begin{aligned} &\lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} [f(\mathbf{p}+\mathbf{t}) - \phi(\mathbf{t})] = \lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} [\|\mathbf{t}\|_{\mathbb{R}^m}\alpha(\mathbf{t}) + f(\mathbf{p})] \\ \iff &\lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} f(\mathbf{p}+\mathbf{t}) = f(\mathbf{p}). \end{aligned}$$

Thus, f is continuous at $\mathbf{p}$. ∎

Theorem 1.1.3. With the condition in Definition 1.1.1, let $g : \mathbb{R}^n \to \mathbb{R}^k$.

If f is differentiable at $\mathbf{p}$ and g is differentiable at $f(\mathbf{p})$, then $g \circ f$ is differentiable at $\mathbf{p}$.

Proof. As f is differentiable at $\mathbf{p}$, there exists a linear mapping $\phi : \mathbb{R}^m \to \mathbb{R}^n$ and a neighbourhood N of $\mathbf{p}$ such that for any $\mathbf{t} \in \mathbb{R}^m$ with $\mathbf{p}+\mathbf{t} \in \mathbb{R}^m$,

$$f(\mathbf{p}) + \phi(\mathbf{t}) = f(\mathbf{p}+\mathbf{t}) - \|\mathbf{t}\|_{\mathbb{R}^m}\alpha(\mathbf{t}).$$

As g is differentiable at $f(\mathbf{p})$, there exists a linear mapping $\lambda : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} \frac{g(f(\mathbf{p}) + \|\mathbf{t}\|_{\mathbb{R}^m}\phi(\hat{\mathbf{t}})) - f(\mathbf{p}) - \lambda(\|\mathbf{t}\|_{\mathbb{R}^m}\phi(\hat{\mathbf{t}}))}{\left\| \|\mathbf{t}\|_{\mathbb{R}^m}\phi(\hat{\mathbf{t}}) \right\|_{\mathbb{R}^n}} = \mathbf{0}_{\mathbb{R}^k}.$$

As ϕ is linear, we have

$$\|\mathbf{t}\|_{\mathbb{R}^m}\phi(\hat{\mathbf{t}}) = \phi(\mathbf{t}).$$

By scalar multiplication, we have

$$\left\| \|\mathbf{t}\|_{\mathbb{R}^m}\phi(\hat{\mathbf{t}}) \right\|_{\mathbb{R}^n} = \|\mathbf{t}\|_{\mathbb{R}^m}\|\phi(\hat{\mathbf{t}})\|_{\mathbb{R}^n}.$$

Now, we have

$$\begin{aligned} & \lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} \frac{g(f(\mathbf{p})+\phi(\mathbf{t}))-g(f(\mathbf{p}))-\lambda(\phi(\mathbf{t}))}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \mathbf{0}_{\mathbb{R}^k} \\ \iff \quad & \lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} \frac{g(f(\mathbf{p}+\mathbf{t}))-g(f(\mathbf{p}))-(\lambda\circ\phi)(\mathbf{t})}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \mathbf{0}_{\mathbb{R}^k}. \end{aligned}$$

As λ and ϕ are both linear, $\lambda \circ \phi$ are also linear.

By Definition 1.1.1, $g \circ f$ is differentiable at $\mathbf{p}$. ∎

§1.2 Directional Derivatives

Observation 1.2.1. Let $f : \mathbb{R}^m \to \mathbb{R}^n$, and let $g : \mathbb{R} \to \mathbb{R}^m$ be defined as

$$g(t) := \mathbf{p} + t\mathbf{u},$$

where $\mathbf{p}, \mathbf{u} \in \mathbb{R}^m$ and $\mathbf{u} \neq \mathbf{0}_{\mathbb{R}^m}$.

Let $h = f \circ g$ and define $h' : D_{h'} \subseteq \mathbb{R} \to \mathbb{R}^n$ as

$$h'(t) := \lim_{t\to t_0} \frac{h(t)-h(t_0)}{t-t_0},$$

where for any $t \in D_{h'}$, the this limit exists in $\mathbb{R}^n$. Thus,

$$h'(0) = \lim_{t\to 0} \frac{f(\mathbf{p}+t\mathbf{u})-f(\mathbf{p})}{t}$$

describes the instantaneous rate of change of f along the straight line $\{\mathbf{p} + t\mathbf{u} : t \in \mathbb{R}\}$ with $\|\mathbf{u}\|_{\mathbb{R}^m}$ as the unit length. $h'(0)$ is so-called the $\mathbf{u}$-directional derivative of f at $\mathbf{p}$ (See Definition 1.2.1).

Definition 1.2.1 (Directional Derivatives)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^n$, and let $\mathbf{u} \in \mathbb{R}^m \setminus \{\mathbf{0}_{\mathbb{R}^m}\}$. The $\mathbf{u}$-*derived function* of f, denoted $\nabla_{\mathbf{u}} f$ is a function $\nabla_{\mathbf{u}} f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ defined as

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

where D is the set of all $\mathbf{x} \in \mathbb{R}^m$ such that $\nabla_{\mathbf{u}} f(\mathbf{x})$ exists in $\mathbb{R}^n$. Let $\mathbf{p} \in D$, then $\nabla_{\mathbf{u}} f(\mathbf{p})$ is a $\mathbf{u}$-*directional derivative* of f at $\mathbf{p}$.

Note 1.2.1. As $\mathbb{R}$ is an ordered field, there are only two direction in $\mathbb{R}$. Thus, for any $u \in \mathbb{R} \setminus \{0\}$, $u > 0$ or $u < 0$. If $u = 1$, then we write

$$\frac{\mathrm{d}f}{\mathrm{d}t} \text{ or } f' \text{ for } \nabla_u f,$$

and simply call f' the *derived function* of f. If f is differentiable at a point $p \in \mathbb{R}$, then $f'(p)$ is called the *derivative* of f at p.

Theorem 1.2.1. With the condition in Definition 1.2.1, for any $s \in \mathbb{R}\setminus\{0\}$,

$$\nabla_{s\mathbf{u}} f(\mathbf{p}) = s\nabla_{\mathbf{u}} f(\mathbf{p}).$$

Proof. Let $\theta = ts^{-1}$, then, by Definition 1.2.1, we have

$$\begin{aligned} s\nabla_{\mathbf{u}} f(\mathbf{p}) &= s \lim_{t\to 0} \frac{f(\mathbf{p} + t\mathbf{u}) - f(\mathbf{p})}{t} \\ &= \lim_{t\to 0} \frac{f(\mathbf{p} + t\mathbf{u}) - f(\mathbf{p})}{ts^{-1}} \\ &= \lim_{t\to 0} \frac{f(\mathbf{p} + \theta(s\mathbf{u})) - f(\mathbf{p})}{\theta} \\ &= \nabla_{s\mathbf{u}} f(\mathbf{p}). \end{aligned}$$

■

Theorem 1.2.2. With the condition in Definition 1.2.1, if f is differentiable at $\mathbf{p} \in \mathbb{R}^m$, then, in Definition 1.1.1, the linear map ϕ is defined as

$$\phi(\mathbf{u}) := \nabla_{\mathbf{u}} f.$$

Proof. By Definition 1.1.1, as f is differentiable at $\mathbf{p}$, then there exists a linear map $\phi : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\phi(\mathbf{t}) = f(\mathbf{p} + \mathbf{t}) - f(\mathbf{t}) - o(\mathbf{t}) \quad \text{as } \mathbf{t} \to \mathbf{0}_{\mathbb{R}^m}.$$

Let $\mathbf{t} = t\mathbf{u}$, then we have

$$\phi(\mathbf{u}) = \lim_{t\to 0} \frac{f(\mathbf{p} + t\mathbf{u}) - f(\mathbf{p})}{t} = \nabla_{\mathbf{u}} f(\mathbf{p}).$$

■

Theorem 1.2.3. With the condition in Definition 1.2.1, if $\nabla_{\mathbf{u}} f(\mathbf{p})$ exists, then there exists an open subset $U \subseteq \mathbb{R}^m$ with $\mathbf{p} \in U$ such that f is relative continuous on the line described by $\mathbf{p} + t\mathbf{u}$ for some $t \in \mathbb{R}$.

Proof. Let U be an open subset of $\mathbb{R}^m$, and let $g : \mathbb{R} \to \mathbb{R}^m$ be defined as

$$g(t) := \mathbf{p} + t\mathbf{u}.$$

Then f is relative continuous on the line defined by $\mathbf{p} + t\mathbf{u}$ for some $t \in \mathbb{R}$ iff $f \restriction_{g[\mathbb{R}]}$ is continuous on $U \cap g[\mathbb{R}]$.

Let $h = f \circ g$, then

$$\nabla_{\mathbf{u}} f(\mathbf{p}) = \lim_{t \to 0} \frac{h(t) - h(0)}{t} = \mathbf{v} \in \mathbb{R}^n.$$

Then, there exists an $\alpha : \mathbb{R} \to \mathbb{R}^n$ with $\alpha(t) \to \mathbf{0}_{\mathbb{R}^n}$ as $t \to 0$, such that there exists an open subset $I \subseteq \mathbb{R}$ with $0 \in I$, such that for any $t \in I$,

$$h(t) = t\mathbf{v} + t\alpha(t) + h(0).$$

Then we have

$$\begin{aligned} & \lim_{t \to 0} h(t) = \lim_{t \to 0} [t\mathbf{v} + t\alpha(t) + h(0)] \\ \iff & \lim_{t \to 0} h(t) = h(0). \end{aligned}$$

Thus, h is continuous at 0.

As it is easy to show g is bijective, $g \circ g^{-1}$ is an identity mapping on $g[\mathbb{R}] \subseteq \mathbb{R}^m$. As composition of mappings is associative, we have

$$\begin{aligned} h = f \circ g \iff & h \circ g^{-1} = f \circ g \circ g^{-1} \\ \iff & h \circ g^{-1} = f \circ (g \circ g^{-1}) \\ \iff & h \circ g^{-1} = f \restriction_{g[\mathbb{R}]} . \end{aligned}$$

It is also easy to find that g^{-1} is continuous everywhere, thus, as h is continuous at 0, $f \restriction_{g[\mathbb{R}]}$ is continuous on $U \cap g[\mathbb{R}]$. Thus, f is relative continuous on the line defined by $\mathbf{p} + t\mathbf{u}$ for some $t \in \mathbb{R}$. ■

Theorem 1.2.4. With the condition in Definition 1.2.1, if f is differentiable at $\mathbf{p}$, then, for any $\mathbf{u} \in \mathbb{R}^m$, $\nabla_{\mathbf{u}} f$ is continuous at $\mathbf{p}$.

Proof. As f is continuous, it is easy to show that

$$\lim_{t \to 0} \nabla_{\mathbf{u}} f(\mathbf{p} + t\mathbf{u}) = \nabla_{\mathbf{u}} f(\mathbf{p}) = \lim_{t \to 0} \frac{f(\mathbf{p} + t\mathbf{u}) - f(\mathbf{p})}{t}.$$

■

§1.3 Mean Value Theorem in Vector Valued Functions

Lemma 1.3.1. Let $f : \mathbb{R} \to \mathbb{R}$, and let $a, b \in \mathbb{R}$ with $a < b$. Suppose f is continuous on $[a, b]$ and differentiable on (a, b), and $0 \notin f'[(a, b)]$.

Then, f is strictly monotonic on $[a, b]$.

Proof. As f is differentiable on (a, b), by Theorem 1.2.4, f' is continuous on (a, b). This implies, if $0 \notin f'[(a, b)]$, then

$$f'[(a, b)] \subseteq \mathbb{R}_{>0} \text{ or } f'[(a, b)] \subseteq \mathbb{R}_{<0}.$$

Let $c \in (a, b)$. As f is differentiable at c, for any

$$f'(c) = \lim_{t \to 0} \frac{f(c + t) - f(c)}{t}.$$

Now, Consider $f'(c) > 0$. Then $f(c + t) - f(c) > 0$ as $t \to 0^+$, and $f(c + t) - f(c) < 0$ as $t \to 0^-$. That is, for any $d, e \in (a, b)$,

$$e < c < d \implies f(e) < f(c) < f(d).$$

As f is continuous at a and b, we have

$$\lim_{e \to a} f(e) = f(a) < f(c) < f(b) = \lim_{d \to b} f(d).$$

If $f'(c) < 0$, the proof is similar. ■

Lemma 1.3.2 (Rolle's Theorem)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^n$. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ with $f(\mathbf{a}) = f(\mathbf{b})$. Suppose f is relative continuous on $\ell[\mathbf{a}, \mathbf{b}]$, and relative differentiable on $\ell(\mathbf{a}, \mathbf{b})$.

Then, there exits $\mathbf{c} \in \ell(\mathbf{a}, \mathbf{b})$ such that $\nabla_{\mathbf{u}} f(\mathbf{c}) = \mathbf{0}_{\mathbb{R}^n}$, where $\mathbf{u} = \mathbf{b} - \mathbf{a}$.

Proof. First, consider $f = \langle f_i \rangle_{i=1}^n$.

Suppose for any $\mathbf{c} \in \ell(\mathbf{a}, \mathbf{b})$, $\nabla_{\mathbf{u}} f(\mathbf{c}) \neq \mathbf{0}_{\mathbb{R}^n}$, then there exists $i \in \{1, \ldots, n\}$ such that $\nabla_{\mathbf{u}} f_i(\mathbf{c}) \neq 0$.

Let $g : \mathbb{R} \to \mathbb{R}^m$ be defined as

$$g(t) = \mathbf{b} - t\mathbf{a},$$

and let $h_i = f_i \circ g$. Then, for any $t \in (0, 1)$, $h_i'(t) \neq 0$.

As f_i is differentiable on $g[(0, 1)]$, and g is differentiable on $(0, 1)$, by Theorem 1.1.3, h_i is differentiable on $(0, 1)$. In this case, $0 \notin h_i'[(0, 1)]$ implies h_i is strictly monotonic (Lemma 1.3.1). This implies

$$h_i(0) = f_i(\mathbf{a}) \neq f_i(\mathbf{b}) = h_i(1).$$

As $f(\mathbf{a}) = \langle f_i(\mathbf{a}) \rangle_{i=1}^n$ and $f(\mathbf{b}) = \langle f_i(\mathbf{b}) \rangle_{i=1}^n$, we have $f(\mathbf{a}) \neq f(\mathbf{b})$. This contradicts the assumption that $f(\mathbf{a}) = f(\mathbf{b})$.

Thus, there has to be a $\mathbf{c} \in \ell(\mathbf{a}, \mathbf{b})$ such that $\nabla_{\mathbf{u}} f_i(\mathbf{c})$. ∎

Lemma 1.3.3. Let $f : \mathbb{R} \to \mathbb{R}^n$. If f is differentiable on open subset (a, b), and continuous on closed interval $[a, b]$, then there exists a $c \in I$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof. Let $\phi : \mathbb{R} \to \mathbb{R}^n$ be defined as

$$\phi(t) := t\frac{f(b) - f(a)}{b - a}.$$

Let $h : \mathbb{R} \to \mathbb{R}^n$ be defined as

$$h(t) := f(t) - \phi(t).$$

Then it is easy to find that

$$h(a) = h(b).$$

As f and ϕ are differentiable on (a, b), so is h. (Why?)

As f and ϕ are continuous on $[a, b]$, so is h. (Why?)

Thus, by Lemma 1.3.2, there exists a $c \in (a, b)$ such that we have

$$\begin{aligned} 0 = h'(c) &= f'(c) - \frac{f(b) - f(a)}{b - a} \\ \iff f'(c) &= \frac{f(b) - f(a)}{b - a}. \end{aligned}$$

■

Theorem 1.3.1 (Mean Value Theorem on $\mathbb{R}^m \to \mathbb{R}^n$)**.**

Let $f : \mathbb{R}^m \to \mathbb{R}^n$. Let $\mathbf{p}, \mathbf{q} \in \mathbb{R}^m$, for convenience, let $g : \mathbb{R} \to \mathbb{R}^m$ be defined as

$$g(t) := \mathbf{p} + t(\mathbf{q} - \mathbf{p}).$$

If $f \restriction_{g[\mathbb{R}]}$ is continuous on $g[(0, 1)]$, and differentiable on $g[[0, 1]]$, then

$$\|f(\mathbf{q}) - f(\mathbf{p})\|_{\mathbb{R}^n} \leq \sup_{\mathbf{x} \in g[(a,b)]} \|\nabla_{\mathbf{u}} f(\mathbf{x})\|_{\mathbb{R}^n}.$$

Proof. Let $h = f \circ g$. As f is continuous on $g[(0, 1)]$ and g is continuous everywhere, h is continuous on $(0, 1)$. By Theorem 1.1.3, as f is differentiable on $g[[0, 1]]$ and g is differentiable on $[0, 1]$, then, by Theorem 1.1.3, h is differentiable on $[0, 1]$.

Let $h' : D \subseteq \mathbb{R} \to \mathbb{R}^n$ be defined as

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

where D is the set of all points in $\mathbb{R}$ such that the limit exists in $\mathbb{R}^n$.

By Lemma 1.3.3, there exists a $c \in (0, 1)$ such that

$$h'(c) = \frac{h(1) - h(0)}{1 - 0}.$$

Now, we have

$$\begin{aligned} h'(c) &= \lim_{t\to 0} \frac{h(c+t)-h(c)}{t} \\ &= \lim_{t\to 0} \frac{f(g(c+t))-f(c)}{t} \\ &= \lim_{t\to 0} \frac{f(\mathbf{p}+c\mathbf{u}+t\mathbf{u})-f(\mathbf{p}+c\mathbf{u})}{t}\Bigg|_{\mathbf{u}=\mathbf{q}-\mathbf{p}} \\ &= \lim_{t\to 0} \frac{f(\mathbf{c}+t\mathbf{u})-f(\mathbf{c})}{t}\Bigg|_{\mathbf{c}=\mathbf{p}+c\mathbf{u}} \\ &= \nabla_{\mathbf{u}} f(\mathbf{c}). \end{aligned}$$

Thus, there exists a $\mathbf{c} \in g[(0,1)]$ such that

$$\nabla_{\mathbf{u}} f(\mathbf{c}) = h(1) - h(0) = f(\mathbf{q}) - f(\mathbf{p}).$$

This implies that there exists some $\mathbf{x} \in g[(0,1)]$ such that

$$\|\nabla_{\mathbf{u}} f(\mathbf{x})\| \geq \|\nabla_{\mathbf{u}} f(\mathbf{c})\|.$$

Thus,

$$\|f(\mathbf{q}) - f(\mathbf{p})\| \leq \sup_{\mathbf{x}\in g[(0,1)]} \|\nabla_{\mathbf{u}} f(\mathbf{x})\|.$$

■

§1.4 Partial Derivatives and Jacobian Matrices

Definition 1.4.1 (Partial Derivatives)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^n : \mathbf{x} \mapsto \mathbf{y}$.

The *i-th partial derived function* of f, denoted $\frac{\partial f}{\partial x_i}$, is the $\hat{\mathbf{e}}_i$-directional derived function of f, where $\hat{\mathbf{e}}_i$ denotes the i-th basis of $\mathbb{R}^m$. If $\frac{\partial f}{\partial x_i}(\mathbf{p})$ exists in $\mathbb{R}^n$ for a $\mathbf{p} \in \mathbb{R}^m$, then this value is called *i-th partial derivative* of f at $\mathbf{p}$.

Definition 1.4.2 (Jacobian Matrices)**.** With the condition in Definition 1.4.1, The *Jacobian Matrix* of f is a function $\nabla f : D \subseteq \mathbb{R}^m \to \mathbb{R}^{n\times m}$ be defined as

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_m} \end{bmatrix},$$

where D is the set of all $\mathbf{x} \in \mathbb{R}^m$ such that $\frac{\partial f}{\partial x_i}$ exists in $\mathbb{R}^m$ for any $i \in \{1, \ldots, n\}$.

Note 1.4.1. If f is considered as an $1 \times n$ matrix, then ∇ can be considered as a function from $\mathbb{F}$ to $\mathbb{S}$ where the domain $\mathbb{F}$ is a normed space contains all functions from $\mathbb{R}^m$ to $\mathbb{R}^n$, and the codomain $\mathbb{S}$ is another normed space contains all $n \times m$ matrices. It is defined as

$$\nabla f := \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_m} \end{bmatrix} \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}.$$

In this sense, it is easy to prove that ∇ is linear by matrices multiplication. Also, the $\mathbf{u}$-directional derived function of f can be considered as

$$\nabla_{\mathbf{u}} f = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_m} \end{bmatrix} \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix} = \mathbf{u}^\top \nabla f.$$

For convenience, we denote

$$(\mathbf{u}^\top \nabla)^k f = \mathbf{u}^\top \nabla \Big(\cdots \big(\mathbf{u}^\top \nabla (\mathbf{u}^\top \nabla f)\big) \cdots \Big) \quad k \text{ times}.$$

In the case $f : \mathbb{R}^m \to \mathbb{R}$, as $f(\mathbf{p}) \in \mathbb{R}$ for any $\mathbf{p} \in \mathbb{R}^m$, $\nabla f(\mathbf{p})$ can be considered as an m dimensional vector $(m \times 1)$, which is called *gradient* of f at $\mathbf{p}$. In this case,

$$\mathbf{u} \cdot \nabla f(\mathbf{p}) = \nabla f(\mathbf{p}) \cdot \mathbf{u} = \nabla_{\mathbf{u}} f(\mathbf{p}).$$

where $\cdot$ denotes the inner product.

Theorem 1.4.1 (Chain Rule)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^n : \mathbf{x} \to \mathbf{y}$, and let $g : \mathbb{R}^n \to \mathbb{R}^k : \mathbf{t} \to \mathbf{x}$. For convenience, let $h = g \circ f$.

If f is differentiable at a point $\mathbf{p} \in \mathbb{R}^m$ and g is differentiable at $f(\mathbf{p}) = \mathbf{q} \in \mathbb{R}^n$, then

$$\nabla h(\mathbf{p}) = [\nabla g(f(\mathbf{p}))]^\top \nabla f(\mathbf{p}) \in \mathbb{R}^{k\times m}.$$

Proof. By Theorem 1.1.3, h is differentiable at $\mathbf{p}$, and there exists $\phi : \mathbb{R}^m \to \mathbb{R}^n$ and $\lambda : \mathbb{R}^n \to \mathbb{R}^k$, such that for any $\mathbf{t} \in \mathbb{R}^m$

$$\lim_{\mathbf{t}\to\mathbf{0}_{\mathbb{R}^m}} \frac{g(f(\mathbf{p}) + \phi(\mathbf{t})) - g(f(\mathbf{p})) - \lambda(\phi(\mathbf{t}))}{\|\mathbf{t}\|_{\mathbb{R}^m}} = \mathbf{0}_{\mathbb{R}^k}.$$

■

—

—

—

—

—

—

—

—

—

—

—

—