Differentiation

Lili Zhao

June 14, 2022

For any $f : D_f \subseteq X \to Y$, let X be a normed vector space over $\mathbb{K}_X$ as the scalar field, and let Y be a normed vector space over $\mathbb{K}_Y$ as the scalar field.

Also, for any normed vector space X, denote $\|\cdot\|_X$ for norm function of X.

§1 Differentiable Mappings

Definition 1.1. A mapping $f : D_f \subseteq X \to Y$ is said to be *differentiable* at $\mathbf{p}$ iff there exists a linear mapping $\phi : X \to Y$, such that

$$\phi(\mathbf{u}) \sim f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) \quad : \mathbf{u} \to \mathbf{p}.$$

Proposition 1.1. f is differentiable at a point $\mathbf{p} \in D_f$ iff there exists a linear mapping $\phi : X \to Y$, such that

$$\lim_{\mathbf{u}\to\mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) - \phi(\mathbf{u})}{\|\mathbf{u}\|_X} = \mathbf{0}_Y.$$

Proof. As f is differentiable at $\mathbf{p}$, we have

$$\phi()$$

∎

Proposition 1.2. If f is differentiable, then ϕ in Definition 1.1 is unique.

Proof. Suppose there exists a linear mapping $\lambda : X \to Y$ such that

$$\lim_{\mathbf{u}\to 0_X} \frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) - \lambda(\mathbf{u})}{\|\mathbf{u}\|_X} = \mathbf{0}_Y.$$

It is easy to obtain that

$$\phi(\hat{\mathbf{u}}) = \lambda(\hat{\mathbf{u}}),$$

which implies $\phi = \lambda$. ■

Proposition 1.3. If f is differentiable at $\mathbf{p}$, then f is continuous at $\mathbf{p}$.

Proof. Assume f is differentiable at $\mathbf{p}$, then there exists a linear mapping $\phi : X \to Y$ such that

$$\lim_{\mathbf{u}\to\mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p})}{\|\mathbf{u}\|_X} = \phi(\hat{\mathbf{u}}).$$

Thus, for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any $\mathbf{x} \in B_X(\mathbf{0}_X, \delta) \setminus \{\mathbf{0}_X\}$,

$$\frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p})}{\|\mathbf{u}\|_X} \in B_Y(\phi(\mathbf{u}), \varepsilon) \iff f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) \in B_Y(\phi(\hat{\mathbf{u}}), \varepsilon\|\mathbf{u}\|_X),$$

Note that, as $\mathbf{u} \to \mathbf{0}_X$, we only consider $\|\mathbf{u}\|_X < 0$. So, $\varepsilon\|u\|_X < \varepsilon$.

Thus, we have

$$\begin{aligned}
&\lim_{\mathbf{u}\to\mathbf{0}_X} [f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p})] = \mathbf{0}_Y = f(\mathbf{p}+\mathbf{0}_X) - f(\mathbf{p}) \\
\iff &\lim_{\mathbf{u}\to\mathbf{0}_X} f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) = f(\mathbf{p}) - f(\mathbf{p}) \\
\iff &\lim_{\mathbf{u}\to\mathbf{0}_X} f(\mathbf{p}+\mathbf{u}) = f(\mathbf{p}),
\end{aligned}$$

which implies that f is continuous at $\mathbf{p}$. ■

§2 Asymptotic Notation

Definition 2.1. Let $f : D_f \subseteq X \to Y$, and let $\mathbf{p} \in D_f$. Assume

$$\lim_{\mathbf{x}\to\mathbf{p}} f(\mathbf{x}) \in Y.$$

The *little-o* of f as $\mathbf{x} \to \mathbf{p}$ is a set

$$o(f(\mathbf{x})) = \left\{ g : D_g \subseteq X \to Y_g : \lim_{\mathbf{x}\to\mathbf{p}} \frac{g(\mathbf{x})}{\|f(\mathbf{x})\|_Y} = \mathbf{0}_{Y_g} \right\}, \text{ as } \mathbf{x} \to \mathbf{p},$$

where for any mapping g, Y_g is a normed vector space over $\mathbb{K}_{Y_g}$.

Equivalently, for any mapping g defined on X, $g \in o(f(\mathbf{p}))$ iff for any $\varepsilon \in \mathbb{R}_{>0}$, there exists a neighbourhood N of $\mathbf{p}$, such that for any $\mathbf{x} \in N$,

$$\left\| \frac{g(\mathbf{x})}{\|f(\mathbf{x})\|_Y} \right\|_{Y_g} < \varepsilon,$$

or, equivalently,

$$\frac{g(\mathbf{x})}{\|f(\mathbf{x})\|_Y} \in B(\mathbf{0}_{Y_g}, \varepsilon).$$

Proposition 2.1. A mapping $f : D_f \subseteq X \to Y$ is differentiable at a point $\mathbf{p} \in D_f$ iff there exists a linear mapping $\phi : X \to Y$, such that

$$f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h}) = o(\mathbf{h})$$

as $\mathbf{h} \to \mathbf{0}_X$.

Proof. f is differentiable at $\mathbf{p}$ iff there exists an $\alpha : D_\alpha \subseteq X \to Y$ with $\alpha(\mathbf{x}) \to \mathbf{0}_Y$ as $\mathbf{x} \to \mathbf{0}_X$, such that

$$\begin{aligned}
&\lim_{\mathbf{h}\to\mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h})}{\|\mathbf{h}\|_X} = \lim_{\mathbf{h}\to\mathbf{0}_X} \alpha(\mathbf{h}) \\
\iff &\lim_{\mathbf{h}\to\mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h})}{\|\mathbf{h}\|_X} = \lim_{\mathbf{h}\to\mathbf{0}_X} \frac{\alpha(\mathbf{h})\|\mathbf{h}\|_X}{\|\mathbf{h}\|_X} \\
\iff &\lim_{\mathbf{h}\to\mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h})}{\|\mathbf{h}\|_X} = \lim_{\mathbf{h}\to\mathbf{0}_X} \frac{o(\mathbf{h})}{\|\mathbf{h}\|_X}.
\end{aligned}$$

The equation holds iff for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any $\mathbf{h} \in B(\mathbf{0}_X, \delta) \setminus \{\mathbf{0}_X\}$,

$$\begin{aligned}
&f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h}) - o(\mathbf{h}) \in B(\mathbf{0}_Y, \varepsilon) \\
\iff &f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h}) \in B(o(\mathbf{h}), \varepsilon).
\end{aligned}$$

This holds iff

$$f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - \phi(\mathbf{h}) = o(\mathbf{h}),$$

as $\mathbf{h} \to \mathbf{0}_X$. ■

Definition 2.2. The *big-O* of f as $\mathbf{x} \to \mathbf{p}$ is a set

$$O(f(\mathbf{x})) = \left\{ g : D_g \subseteq X \to Y_g : \lim_{\mathbf{x}\to\mathbf{p}} \frac{g(\mathbf{x})}{\|f(\mathbf{x})\|_Y} \in Y_g \right\}, \text{ as } \mathbf{x} \to \mathbf{p},$$

where for any mapping g, Y_g is a normed vector space over $\mathbb{K}_{Y_g}$.

Equivalently, for any mapping g defined on X, $g \in O(f(\mathbf{p}))$ iff there exists an $\varepsilon \in \mathbb{R}_{>0}$, such that.........

Proposition 2.2. A mapping $f : D_f \subseteq X \to Y$ is differentiable at a point $\mathbf{p} \in D_f$ iff there exists a linear mapping $\phi : X \to Y$, such that

$$f(\mathbf{p}+\mathbf{h}) \sim l(\mathbf{h}), \text{ as } \mathbf{h} \to \mathbf{0}_X,$$

where

$$l(\mathbf{h}) = f(\mathbf{p}) + \phi(\mathbf{h}).$$

Proof.

$$f(\mathbf{p}+\mathbf{h}) = f(\mathbf{p}) + \phi(\mathbf{h}) + o(\mathbf{h}).$$

■

§3 Directional Derivatives

Definition 3.1. Let $f : D_f \subseteq X \to Y$. Let $\mathbf{u} \in X$ and let $\mathbf{p} \in D_f$.

The $\mathbf{u}$-*directional derivative* of f at $\mathbf{p}$ is defined as

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

if the limit exists in Y.

Proposition 3.1. If $\nabla_{\mathbf{u}} f(\mathbf{p}) \in Y$, then

$$\nabla_{\mathbf{u}} f(\mathbf{p}) = \left.\frac{df(r(t))}{dt}\right|_{t=0},$$

where $r : \mathbb{R} \to X : t \mapsto \mathbf{p} + t\mathbf{u}$.

Proof. Let $r : \mathbb{R} \to X$ be defined as

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

Let, $h = f \circ r$.

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

■

Proposition 3.2. If f is differentiable at p, then $\nabla_{\mathbf{u}} f(\mathbf{p}) \in Y$ for any $\mathbf{u} \in X$.

In particular, the linear mapping ϕ in Definition 1.1 is defined as

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

Proof. Assume f is differentiable at $\mathbf{p}$, then there exists a linear mapping $\phi : X \to Y$ such that

$$\lim_{\mathbf{u} \to \mathbf{0}_X} \frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p})}{\|\mathbf{u}\|_X} = \phi\left(\frac{\mathbf{u}}{\|\mathbf{u}\|_X}\right).$$

Let $\mathbf{u} = t\mathbf{w}$ where $t \in \mathbb{R}$ and $\mathbf{w} \in X$. Then, we have

$$\begin{aligned} &\lim_{t \to 0} \frac{f(\mathbf{p}+t\mathbf{w}) - f(\mathbf{p})}{|t|\|\mathbf{u}\|_X} = \phi\left(\frac{t}{|t|}\mathbf{w}\right) \\ \iff &\lim_{t \to 0} \frac{f(\mathbf{p}+t\mathbf{w}) - f(\mathbf{p})}{t} = \phi(\mathbf{w}) = \nabla_{\mathbf{w}} f(\mathbf{p}). \end{aligned}$$

■

Proposition 3.3. If $\nabla_{\mathbf{u}} f(\mathbf{p}) \in Y$ for a given $\mathbf{u} \in X \setminus \{\mathbf{0}_X\}$, then

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

for any $s \in \mathbb{R}$.

If f is differentiable at $\mathbf{p}$, then, for any $\mathbf{u}, \mathbf{v} \in X$.

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

Proof. Let $t = s\theta$, then we have

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

Assume f is differentiable at $\mathbf{p}$. Let $\phi : X \to Y$ be defined as

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

By Proposition 3.2, ϕ is linear, so for any $u, v \in X$,

$$\nabla_{\mathbf{u}+\mathbf{v}} f(\mathbf{p}) = \phi(\mathbf{u}+\mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v}) = \nabla_{\mathbf{u}} f(\mathbf{p}) + \nabla_{\mathbf{v}} f(\mathbf{p}).$$

■

Example 3.1. However, f is not necessarily differentiable at $\mathbf{p}$, even if $\nabla_{\mathbf{u}} f(\mathbf{p}) \in Y$ for any $\mathbf{u} \in X \setminus \{\mathbf{0}_X\}$.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(x, y) := \begin{cases} 0 & : x \neq y, \\ x + y & : x = y. \end{cases}$$

For any $(u_x, u_y) \in \mathbb{R}^2$, $f(u_x, u_y) \in \mathbb{R}$.

Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$\varphi(x, y) := \nabla_{(x,y)} f(0, 0).$$

If f is differentiable, φ should be linear. But,

$$2 = \varphi(1, 1) \neq \varphi(1, 0) + \varphi(0, 1) = 0.$$

Thus, φ is not linear.

§4 Partial Derivatives

Definition 4.1. Let $f : D_f \subseteq X \to Y$, where $X = \mathbb{K}^n$. Let $\mathbf{p} \in X$

The *partial derivative* of f at $\mathbf{p}$ respect to x_i ($i \in \{1, \ldots, n\}$) is defined as the $\hat{\mathbf{e}}_i$-directional derivative of f at $\mathbf{p}$; i.e.,

$$\frac{\partial f}{\partial x_i}(\mathbf{p}) := \nabla_{\hat{\mathbf{e}}_i} f(\mathbf{p}) = \lim_{t \to 0} \frac{f(\mathbf{p} + t\hat{\mathbf{e}}_i) - f(\mathbf{p})}{t}.$$

Proposition 4.1 (symmetry of second derivatives)**.** If $\dfrac{\partial f}{\partial x_i \partial x_k}$ and $\dfrac{\partial f}{\partial x_k \partial x_i}$ both exist in Y for some $i, j \in \{1, \ldots, n\}$, then

$$\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{p}) = \frac{\partial^2 f}{\partial x_k \partial x_i}.$$

Proof.

$$\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{p}) = \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_k}(\mathbf{p}) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial x_k}(\mathbf{p} + t\hat{\mathbf{e}}_i) - \frac{\partial f}{\partial x_k}(\mathbf{p})}{t}. \tag{i}$$

Consider

$$\begin{aligned}
&\frac{\partial f}{\partial x_k}(\mathbf{p}+t\hat{\mathbf{e}}_i) - \frac{\partial f}{\partial x_k}(\mathbf{p}) \\
=& \lim_{t\to 0} \frac{f(\mathbf{p}+t\hat{\mathbf{e}}_i+t\hat{\mathbf{e}}_k) - f(\mathbf{p}+t\hat{\mathbf{e}}_i)}{t} - \lim_{t\to 0} \frac{f(\mathbf{p}+t\hat{\mathbf{e}}_k) - f(\mathbf{p})}{t} \\
=& \lim_{t\to 0} \frac{f(\mathbf{p}+t\hat{\mathbf{e}}_k+t\hat{\mathbf{e}}_i) - f(\mathbf{p}+t\hat{\mathbf{e}}_k)}{t} - \lim_{t\to 0} \frac{f(\mathbf{p}+t\hat{\mathbf{e}}_i) - f(\mathbf{p})}{t} \\
=& \frac{\partial f}{\partial x_i}(\mathbf{p}+t\hat{\mathbf{e}}_k) - \frac{\partial f}{\partial x_i}(\mathbf{p}).
\end{aligned} \tag{ii}$$

Substitute (ii) into (i), we have

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{p}) &= \lim_{t\to 0} \frac{\frac{\partial f}{\partial x_i} f(\mathbf{p}+t\hat{\mathbf{e}}_k) - \frac{\partial f}{\partial x_i} f(\mathbf{p})}{t} \\
&= \frac{\partial}{\partial x_k} \frac{\partial f}{\partial x_i}(\mathbf{p}) \\
&= \frac{\partial^2 f}{\partial x_k \partial x_i}.
\end{aligned}$$

■

§5 Jacobian Matrices and Gradient

Definition 5.1. Let $f : D_f \subseteq X \to Y$, where $X = \mathbb{K}_X^n$.

The *Jacobian matrix* of f at $\mathbf{p}$ is defined as

$$\mathbf{J}_f(\mathbf{p}) := \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{p}) & \dots & \frac{\partial f}{\partial x_n}(\mathbf{p}) \end{bmatrix}.$$

Note 5.1. Customarily, if $Y = \mathbb{K}_Y$, we call $\mathbf{J}_f^\top(\mathbf{p})$ the *gradient* of f at $\mathbf{p}$, and denote $\nabla f(\mathbf{p})$ for $\mathbf{J}_f^\top(\mathbf{p})$. If $Y = \mathbb{K}_Y^m$, then

$$\mathbf{J}_f(\mathbf{p}) = \begin{bmatrix} \nabla^\top f_1(\mathbf{p}) \\ \vdots \\ \nabla^\top f_m(\mathbf{p}) \end{bmatrix}.$$

Proposition 5.1. Let $f : D_f \subseteq X \to Y$, where $X = \mathbb{K}_X^n$. If f is differentiable at $\mathbf{p}$, then we have.

$$\nabla_{\mathbf{u}} f(\mathbf{p}) = \mathbf{J}_f(\mathbf{p})\,\mathbf{u}.$$

Proposition 5.2. Let $X = \mathbb{K}_X^m$, $Y = \mathbb{K}_Y^n$.

Let $f : D_f \subseteq X \to Y$ be differentiable at $\mathbf{p}$.

Let $g : D_g \subseteq Y \to Z$ be differentiable at $f(\mathbf{p})$.

For any $i \in \{1, \dots, m\}$, we have

$$\frac{\partial(g \circ f)}{\partial x_i}(\mathbf{p}) = \mathbf{J}_g(f(\mathbf{p}))\frac{\partial f}{\partial x_i}(\mathbf{p}).$$

Proof. Define $\varphi(t) := f(\mathbf{p} + t\hat{\mathbf{e}}_i)$, where $\hat{\mathbf{e}}_i$ denotes the i-th basis of X. Then,

$$\frac{\partial(g \circ f)}{\partial x_i}(\mathbf{p}) = \lim_{t \to 0} \frac{g(\varphi(t)) - g(\varphi(0))}{t}.$$

Assume there exists neighbourhood N of $\mathbf{p}$ such that $f \restriction_N$ is constant, then the limit above is zero, and $\frac{\partial f}{\partial x_i}(\mathbf{p})$ is also zero. There is nothing to prove in this case. So, assume that for any neighbourhood N of $\mathbf{p}$, $f \restriction_N$ is not constant.

As f is differentiable at $\mathbf{p}$, $\frac{\partial f}{\partial x_i}(\mathbf{p}) \in Y$, and $\varphi'(0) \in Y$. (This is also a chain rule, but why?) So, we can define

$$\lambda(t) := t\varphi'(0) + \varphi(0).$$

As

$$\lim_{h \to 0} \frac{\varphi(h)}{\|\lambda(h)\|_Y} = \mathbf{1}_Y = \lim_{h \to 0} \frac{\lambda(h)}{\|\varphi(h)\|_Y},$$

we have $\varphi \to \lambda$ as $h \to 0$. Thus, we have

$$\begin{aligned}
\frac{\partial(g \circ f)}{\partial x_i} &= \lim_{t \to 0} \frac{g(\varphi(0) + t\varphi'(0)) - g(\varphi(0))}{t} \\
&= \nabla_{\varphi'(0)} g(\varphi(0)) \\
&= \mathbf{J}_g(\varphi(0))\, \varphi'(0) \\
&= \mathbf{J}_g(f(\mathbf{p}))\, \frac{\partial f}{\partial x_i}(\mathbf{p}).
\end{aligned}$$

■

Proposition 5.3. With the conditions above, we have

$$\mathbf{J}_{g \circ f}(\mathbf{p}) = \mathbf{J}_g(f(\mathbf{p}))\, \mathbf{J}_f(\mathbf{p}).$$

Proof. First, consider $\mathbf{J}_{g \circ f}(\mathbf{p})$ as an $1 \times m$ matrix:

$$\mathbf{J}_{g \circ f}(\mathbf{p}) = \begin{bmatrix} \frac{\partial(g \circ f)}{\partial x_1}(\mathbf{p}) & \cdots & \frac{\partial(g \circ f)}{\partial x_m}(\mathbf{p}) \end{bmatrix}.$$

For any $i \in \{1, \ldots, m\}$, we have

$$\frac{\partial (g \circ f)}{\partial x_i}(\mathbf{p}) = \mathbf{J}_g(f(\mathbf{p}))\, \frac{\partial f}{\partial x_i}(\mathbf{p}).$$

Then we have

$$\mathbf{J}_{g\circ f}(\mathbf{p}) = \mathbf{J}_g(f(\mathbf{p})) \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{p}) & \cdots & \dfrac{\partial f}{\partial x_m}(\mathbf{p}) \end{bmatrix} = \mathbf{J}_g(f(\mathbf{p}))\, \mathbf{J}_f(\mathbf{p}).$$

The proof is done. ■

Note 5.2. In this proof, if $Z = \mathbb{K}_Z^r$, then,

$$\begin{aligned} \mathbf{J}_{g\circ f}(\mathbf{p}) &= \begin{bmatrix} \nabla^\top g_1(f(\mathbf{p}))\, \dfrac{\partial f}{\partial x_1}(\mathbf{p}) & \cdots & \nabla^\top g_1(f(\mathbf{p}))\, \dfrac{\partial f}{\partial x_m}(\mathbf{p}) \\ \vdots & \ddots & \vdots \\ \nabla^\top g_r(f(\mathbf{p}))\, \dfrac{\partial f}{\partial x_1}(\mathbf{p}) & \cdots & \nabla^\top g_r(f(\mathbf{p}))\, \dfrac{\partial f}{\partial x_m}(\mathbf{p}) \end{bmatrix} \\ &= \begin{bmatrix} \nabla^\top g_1(f(\mathbf{p})) \\ \vdots \\ \nabla^\top g_r(f(\mathbf{p})) \end{bmatrix} \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{p}) & \cdots & \dfrac{\partial f}{\partial x_m}(\mathbf{p}) \end{bmatrix} \\ &= \mathbf{J}_g(f(\mathbf{p}))\, \mathbf{J}_f(\mathbf{p}). \end{aligned}$$

Note 5.3. The notation $df(\mathbf{p}, \cdot)$ in *Mathematical Analysis* by Elias Zakon can be considered as a mapping from $X \to Y$ be defined as

$$df(\mathbf{p}, \mathbf{x}) := \nabla_{\mathbf{x}} f(\mathbf{p}) = \mathbf{J}_f(\mathbf{p})\, \mathbf{x}.$$

In this sense, we can consider Proposition 5.3 as

$$\mathbf{J}_{g\circ f}(\mathbf{p}) = \mathbf{J}_g(f(\mathbf{p}))\, \mathbf{J}_f(\mathbf{p}) = dg(f(\mathbf{p}), \cdot) \circ df(\mathbf{p}, \cdot).$$

Briefly "The differential of the composite is the composite of differentials."

§6 Taylor's Theorem

Proposition 6.1 (Taylor's Theorem)**.** Let $f : D_f \subseteq X \to Y$, where $X = \mathbb{K}_X^n$.

If f is C^{k+1} at $\mathbf{p}$, then, there exists an open set $U \subseteq D_f$ with $\mathbf{p} \in U$, such that for any $\mathbf{u} \in U$,

$$f(\mathbf{p}+\mathbf{u}) = \sum_{i=1}^{k} \frac{1}{i!} (\mathbf{J} \times \mathbf{u})_f^i(\mathbf{p}) + R_k,$$

where

$$R_k = \frac{(\mathbf{J} \times \mathbf{u})_f^{k+1}(\mathbf{c})}{(k+1)!}$$

for some $\mathbf{c} \in U$.

Proof. Let $\mathbf{u} \in U$, and let $s \in \mathbb{K}_X$ with $\mathbf{u} = t\hat{\mathbf{u}}$.

Assume f is C^1 at $\mathbf{p}$, then

$$\lim_{t \to 0} \frac{f(\mathbf{p}+\mathbf{u}) - f(\mathbf{p}) - \nabla_{\mathbf{u}} f(\mathbf{p})}{t} = \lim_{t \to 0} \frac{t\alpha(t)}{t} \tag{1}$$

for a mapping $\alpha : \mathbb{K}_X \to Y$ with

$$\lim_{t \to 0} \alpha(t) = 0.$$

Multiply both sides by t (because little-o?)

... ■