Advanced Calculus (II)

WEN-CHING LIEN

Department of Mathematics
National Cheng Kung University

2009

11.4: Chain Rule

Theorem (11.28 Chain Rule)

Suppose that $\mathbf{a} \in \mathbf{R}^n$, *that* g *is a vector function from* n *variables to* m *variables, and that* f *is a vector function from* m *variables to* p *variables. If* g *is differentiable at* $\mathbf{a}$ *and* f *is differentiable at* $g(\mathbf{a})$, *then* $f \circ g$ *is differentiable at* $\mathbf{a}$ *and*

$$D(f \circ g)(\mathbf{a}) = Df(g(\mathbf{a}))Dg(\mathbf{a}). \tag{20}$$

(The product $Df(g(\mathbf{a}))Dg(\mathbf{a})$ *is a matrix multiplication.)*

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

Set $T = Df(g(\mathbf{a}))Dg(\mathbf{a})$ and observe that T, the product of a $p \times m$ matrix with an $m \times n$ matrix, is a $p \times n$ matrix, the right size for the total derivative of $f \circ g$. By the uniqueness of the total derivative, we must show that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{21}$$

Let $\mathbf{b} = g(\mathbf{a})$. Set

$$\varepsilon(\mathbf{h}) = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}) - Dg(\mathbf{a})(\mathbf{h}) \tag{22}$$

and

$$\delta(\mathbf{k}) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) - Df(\mathbf{b})(\mathbf{k}) \tag{23}$$

for $\mathbf{h}$ and $\mathbf{k}$ sufficiently small.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})) + \delta(\mathbf{k}) = T(\mathbf{h}) + Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})) + \delta(\mathbf{k}) = T(\mathbf{h}) + Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})) + \delta(\mathbf{k}) = T(\mathbf{h}) + Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

By hypothesis, $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, and $\frac{\delta(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ in $\mathbf{R}^p$ as $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$. Fix $\mathbf{h}$ small and set $\mathbf{k} = g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})$. Since (23) and (22) imply

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) = f(\mathbf{b}+\mathbf{k}) - f(\mathbf{b}) = Df(\mathbf{b})(\mathbf{k}) + \delta(\mathbf{k})$$

$$= Df(\mathbf{b})(Dg(\mathbf{a})(\mathbf{h})+\varepsilon(\mathbf{h}))+\delta(\mathbf{k}) = T(\mathbf{h})+Df(\mathbf{b})(\varepsilon(\mathbf{h}))+\delta(\mathbf{k}),$$

we have

$$f(g(\mathbf{a}+\mathbf{h})) - f(g(\mathbf{a})) - T(\mathbf{h}) = Df(\mathbf{b})(\varepsilon(\mathbf{h})) + \delta(\mathbf{k})$$

$$=: T_1(\mathbf{h}) + T_2(\mathbf{h}).$$

It remains to verify that $\frac{T_j(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ for $j = 1, 2$.

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm, we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\le \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

□

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm, we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\leq \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

□

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm. we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\leq \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm. we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\leq \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

□

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm. we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\leq \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

□

Proof.

Since $\frac{\varepsilon(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $Df(\mathbf{b})(\mathbf{h})$ is matrix multiplication, it is clear that $\frac{T_1(\mathbf{h})}{\|\mathbf{h}\|} \to Df(\mathbf{b})(\mathbf{0}) = \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. On the other hand, by (22), the triangle inequality, and the definition of the operator norm. we have

$$\|\mathbf{k}\| := \|g(\mathbf{a}+\mathbf{h}) - g(\mathbf{a})\| = \|Dg(\mathbf{a})(\mathbf{h}) + \varepsilon(\mathbf{h})\|$$

$$\leq \|Dg(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|\varepsilon(\mathbf{h})\|.$$

□

Proof.

Thus $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is bounded for $\mathbf{h}$ sufficiently small. Since $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, it follows that

$$\frac{\|T_2(\mathbf{h})\|}{\|\mathbf{h}\|} = \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} \cdot \frac{\|\delta(\mathbf{k})\|}{\|\mathbf{k}\|} \to \mathbf{0}$$

as $\mathbf{h} \to \mathbf{0}$. We conclude that $f \circ g$ is differentiable at $\mathbf{a}$ and the differentiable at $\mathbf{a}$ and the derivative is $Df(g(\mathbf{a}))Dg(\mathbf{a})$. □

Proof.

Thus $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is bounded for $\mathbf{h}$ sufficiently small. Since $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, it follows that

$$\frac{\|T_2(\mathbf{h})\|}{\|\mathbf{h}\|} = \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} \cdot \frac{\|\delta(\mathbf{k})\|}{\|\mathbf{k}\|} \to \mathbf{0}$$

as $\mathbf{h} \to \mathbf{0}$. We conclude that $f \circ g$ is differentiable at $\mathbf{a}$ and the differentiable at $\mathbf{a}$ and the derivative is $Df(g(\mathbf{a}))Dg(\mathbf{a})$. □

Proof.

Thus $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is bounded for $\mathbf{h}$ sufficiently small. Since $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, it follows that

$$\frac{\|T_2(\mathbf{h})\|}{\|\mathbf{h}\|} = \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} \cdot \frac{\|\delta(\mathbf{k})\|}{\|\mathbf{k}\|} \to \mathbf{0}$$

as $\mathbf{h} \to \mathbf{0}$. We conclude that $f \circ g$ is differentiable at $\mathbf{a}$ and the differentiable at $\mathbf{a}$ and the derivative is $Df(g(\mathbf{a}))Dg(\mathbf{a})$.

□

Proof.

Thus $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is bounded for $\mathbf{h}$ sufficiently small. Since $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, it follows that

$$\frac{\|T_2(\mathbf{h})\|}{\|\mathbf{h}\|} = \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} \cdot \frac{\|\delta(\mathbf{k})\|}{\|\mathbf{k}\|} \to \mathbf{0}$$

as $\mathbf{h} \to \mathbf{0}$. We conclude that $f \circ g$ is differentiable at $\mathbf{a}$ and the differentiable at $\mathbf{a}$ and the derivative is $Df(g(\mathbf{a}))Dg(\mathbf{a})$.

Proof.

Thus $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is bounded for $\mathbf{h}$ sufficiently small. Since $\mathbf{k} \to \mathbf{0}$ in $\mathbf{R}^m$ as $\mathbf{h} \to \mathbf{0}$ in $\mathbf{R}^n$, it follows that

$$\frac{\|T_2(\mathbf{h})\|}{\|\mathbf{h}\|} = \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} \cdot \frac{\|\delta(\mathbf{k})\|}{\|\mathbf{k}\|} \to \mathbf{0}$$

as $\mathbf{h} \to \mathbf{0}$. We conclude that $f \circ g$ is differentiable at $\mathbf{a}$ and the differentiable at $\mathbf{a}$ and the derivative is $Df(g(\mathbf{a}))Dg(\mathbf{a})$. □

Example (ex10, p.352)

Let $f : \mathbf{R}^2 \to \mathbf{R}$ be $\mathcal{C}^2$ on $\mathbf{R}^2$ and set $u(r, \theta) = f(r\cos\theta, r\sin\theta)$. If f satisfies the *Laplace equation*, i.e., if

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$$

prove for each $r \neq 0$ that

$$\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial r^2} = 0.$$

Thank you.