MAU23203—Analysis in Several Variables
School of Mathematics, Trinity College
Michaelmas Term 2021
Video Presentation concerning the Multivariable Chain Rule

Trinity College Dublin

Definition

Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. The function φ is said to be *differentiable* at $\mathbf{p}$, with *derivative* $T\colon \mathbb{R}^m \to \mathbb{R}^n$ if and only if

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x}-\mathbf{p})\right) = \mathbf{0}.$$

Definition

Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. The function φ is said to be *differentiable* at $\mathbf{p}$, with *derivative* $T\colon \mathbb{R}^m \to \mathbb{R}^n$ if and only if

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x}-\mathbf{p})\right) = \mathbf{0}.$$

Henceforth we shall usually denote the derivative of a differentiable map $\varphi\colon X \to \mathbb{R}^n$ at a point $\mathbf{p}$ of its domain X by $(D\varphi)_{\mathbf{p}}$.

Lemma A

Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m \to \mathbb{R}^n$ and let $\mathbf{p}$ be a point belonging to the domain X of the function φ. Also let $\sigma\colon X \to \mathbb{R}^n$ be the function defined throughout the domain X of the function φ that is uniquely characterized by the properties that $\sigma(\mathbf{p}) = \mathbf{0}$ and

$$\varphi(\mathbf{x}) = \varphi(\mathbf{p}) + T(\mathbf{x} - \mathbf{p}) + |\mathbf{x} - \mathbf{p}|\, \sigma(\mathbf{x})$$

for all points $\mathbf{x}$ of the domain X of the function φ. Then the function $\varphi\colon X \to \mathbb{R}^n$ is differentiable at the point $\mathbf{p}$, with derivative $T\colon \mathbb{R}^m \to \mathbb{R}^n$, if and only if the associated function σ is continuous at the point $\mathbf{p}$.

Proof

Note that

$$\sigma(\mathbf{x}) = \begin{cases} \dfrac{1}{|\mathbf{x} - \mathbf{p}|} \left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})\right) & \text{if } \mathbf{x} \neq \mathbf{p}; \\ \mathbf{0} & \text{if } \mathbf{x} = \mathbf{p}. \end{cases}$$

The very definition of differentiability therefore ensures that the function φ is differentiable at the point $\mathbf{p}$, with derivative T, if and only if

$$\lim_{\mathbf{x} \to \mathbf{p}} \sigma(\mathbf{x}) = \mathbf{0} = \sigma(\mathbf{p}).$$

Moreover $\lim_{\mathbf{x} \to \mathbf{p}} \sigma(\mathbf{x}) = \sigma(\mathbf{p})$ if and only if the function σ is continuous at the point $\mathbf{p}$. The result follows. ∎

Lemma B

Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Then the function φ is differentiable at $\mathbf{p}$, with derivative T, if and only if, given any positive real number ε, there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| \leq \varepsilon|\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$.

Proof

First suppose that the function $\varphi\colon X \to \mathbb{R}^n$ has the property that, given any positive real number ε_0, there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| \leq \varepsilon_0 |\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$.

Proof

First suppose that the function $\varphi\colon X \to \mathbb{R}^n$ has the property that, given any positive real number ε_0, there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| \leq \varepsilon_0 |\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. Let some positive number ε be given, and let ε_0 be chosen so that $0 < \varepsilon_0 < \varepsilon$.

Proof

First suppose that the function $\varphi\colon X \to \mathbb{R}^n$ has the property that, given any positive real number ε_0, there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| \leq \varepsilon_0 |\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. Let some positive number ε be given, and let ε_0 be chosen so that $0 < \varepsilon_0 < \varepsilon$. Then there exists some positive real number δ such that the above inequality holds at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$.

But then

$$\frac{1}{|\mathbf{x}-\mathbf{p}|}\,|\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})|<\varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0<|\mathbf{x}-\mathbf{p}|<\delta$, But then

$$\frac{1}{|\mathbf{x} - \mathbf{p}|} \left| \varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p}) \right| < \varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0 < |\mathbf{x} - \mathbf{p}| < \delta$, and therefore

$$\lim_{\mathbf{x} \to \mathbf{p}} \frac{1}{|\mathbf{x} - \mathbf{p}|} \left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p}) \right) = \mathbf{0}.$$

Thus the function φ is differentiable at the point $\mathbf{p}$.

Conversely suppose that the function φ is differentiable at the point $\mathbf{p}$. Let some positive real number ε be given. Then there exists some positive real number δ such that

$$\frac{1}{|\mathbf{x}-\mathbf{p}|}\,|\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})|<\varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0<|\mathbf{x}-\mathbf{p}|<\delta$.

Conversely suppose that the function φ is differentiable at the point $\mathbf{p}$. Let some positive real number ε be given. Then there exists some positive real number δ such that

$$\frac{1}{|\mathbf{x}-\mathbf{p}|}\,|\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})|<\varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0<|\mathbf{x}-\mathbf{p}|<\delta$. Considering separately the cases when $\mathbf{x}=\mathbf{p}$ and when $0<|\mathbf{x}-\mathbf{p}|<\delta$, it then follows that

$$|\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})|\leq\varepsilon|\mathbf{x}-\mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x}-\mathbf{p}|<\delta$. The result follows. ∎

Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let S and T be linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. We claim that if both

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - S(\mathbf{x}-\mathbf{p})\right) = \mathbf{0}$$

and

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x}-\mathbf{p})\right) = \mathbf{0},$$

then $S = T$.

Indeed these two conditions taken together would ensure that

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left((T-S)(\mathbf{x}-\mathbf{p})\right) \\
&\quad = \lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x})-\varphi(\mathbf{p})-S(\mathbf{x}-\mathbf{p})\right) \\
&\qquad\quad - \lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})\right) \\
&\quad = \mathbf{0}.
\end{aligned}$$

Indeed these two conditions taken together would ensure that

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left((T-S)(\mathbf{x}-\mathbf{p})\right) \\
&\quad = \lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x})-\varphi(\mathbf{p})-S(\mathbf{x}-\mathbf{p})\right) \\
&\qquad\quad - \lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\left(\varphi(\mathbf{x})-\varphi(\mathbf{p})-T(\mathbf{x}-\mathbf{p})\right) \\
&\quad = \mathbf{0}.
\end{aligned}$$

Therefore, taking a fixed non-zero vector $\mathbf{w}$ in $\mathbb{R}^m$, and setting $\mathbf{x} = \mathbf{p} + t\mathbf{w}$, we find that

$$\lim_{t\to 0}\left(\frac{1}{t|\mathbf{w}|}(T-S)(t\mathbf{w})\right) = \mathbf{0}.$$

The linearity of $T - S$ then ensures that $(T - S)\mathbf{w} = \mathbf{0}$.

We have now shown that $(T - S)\mathbf{w} = \mathbf{0}$ for all non-zero vectors $\mathbf{w}$ in $\mathbb{R}^m$. It follows that $S = T$ as claimed.

We conclude therefore that, if a vector-valued function $\varphi\colon X \to \mathbb{R}^m$ defined over an open set X in $\mathbb{R}^m$ is differentiable at some point $\mathbf{p}$ of the domain X of the function φ, then the derivative $(D\varphi)_{\mathbf{p}}$ of the function φ at the point $\mathbf{p}$ is a linear transformation that is uniquely determined by the function φ and the point $\mathbf{p}$ at which the derivative is to be taken.

Definition

Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The *operator norm* $\|T\|_{\mathrm{op}}$ of T is the smallest non-negative real number with the property that $|T\mathbf{w}| \leq \|T\|_{\mathrm{op}} |\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^m$.

Definition

Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The *operator norm* $\|T\|_{\mathrm{op}}$ of T is the smallest non-negative real number with the property that $|T\mathbf{w}| \leq \|T\|_{\mathrm{op}} |\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^m$.

The operator norm $\|T\|_{\mathrm{op}}$ of a linear transformation $T\colon \mathbb{R}^m \to \mathbb{R}^n$ may be characterized as the maximum value attained by $|T\mathbf{w}|$ as $\mathbf{w}$ ranges over all vectors in $\mathbb{R}^m$ that satisfy $|\mathbf{w}| = 1$.

Proposition C

Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $\mathbf{p}$ be a point of X at which the function φ is differentiable, and let M be a real number satisfying $M > \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$, where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the derivative $(D\varphi)_{\mathbf{p}}$ of φ at $\mathbf{p}$. Then there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M\,|\mathbf{x} - \mathbf{p}|$$

for all points $\mathbf{x}$ of X satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

Proof

Let $\varepsilon = M - \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$. Then $\varepsilon > 0$.

Proof
Let $\varepsilon = M - \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$. Then $\varepsilon > 0$. Now

$$|(D\varphi)_{\mathbf{p}}\mathbf{w}| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}\,|\mathbf{w}|$$

for all $\mathbf{w} \in \mathbb{R}^m$.

Proof

Let $\varepsilon = M - \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$. Then $\varepsilon > 0$. Now

$$|(D\varphi)_{\mathbf{p}}\mathbf{w}| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}\,|\mathbf{w}|$$

for all $\mathbf{w} \in \mathbb{R}^m$. Also the differentiability of the function φ at the point $\mathbf{p}$ ensures that there exists some positive real number δ that is small enough to ensure that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \varepsilon\,|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$ (see Lemma B).

It then follows from the Triangle Inequality satisfied by the Euclidean distance function that

$$\begin{aligned}
&|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \\
&\quad \leq \quad |(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| + |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \\
&\quad \leq \quad |(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| + \varepsilon|\mathbf{x} - \mathbf{p}|
\end{aligned}$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

But the definition of the operator norm ensures that

$$|(D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}\, |\mathbf{x}-\mathbf{p}|$$

for all $\mathbf{x} \in X$.

But the definition of the operator norm ensures that

$$|(D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}\, |\mathbf{x}-\mathbf{p}|$$

for all $\mathbf{x} \in X$. Moreover the value of the positive real number ε has been chosen so as to ensure that $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}} + \varepsilon = M$.

But the definition of the operator norm ensures that

$$|(D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}\,|\mathbf{x}-\mathbf{p}|$$

for all $\mathbf{x} \in X$. Moreover the value of the positive real number ε has been chosen so as to ensure that $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}} + \varepsilon = M$. It follows that

$$\begin{aligned} |\varphi(\mathbf{x}) - \varphi(\mathbf{p})| &\leq |(D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| + \varepsilon|\mathbf{x}-\mathbf{p}| \\ &\leq (\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}} + \varepsilon)|\mathbf{x}-\mathbf{p}| = M\,|\mathbf{x}-\mathbf{p}| \end{aligned}$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x}-\mathbf{p}| < \delta$, as required. ∎

Proposition D (Chain Rule)

Let X and Y be open sets in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, let $\varphi\colon X \to \mathbb{R}^n$ and $\psi\colon Y \to \mathbb{R}^k$ be functions mapping X and Y into $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, where $\varphi(X) \subset Y$, and let $\mathbf{p}$ be a point of X. Suppose that φ is differentiable at $\mathbf{p}$ and that ψ is differentiable at $\varphi(\mathbf{p})$. Then the composition $\psi \circ \varphi\colon X \to \mathbb{R}^k$ is differentiable at $\mathbf{p}$, and

$$D(\psi \circ \varphi)_{\mathbf{p}} = (D\psi)_{\varphi(\mathbf{p})} \circ (D\varphi)_{\mathbf{p}}.$$

Thus the derivative of the composition $\psi \circ \varphi$ of the functions at the point $\mathbf{p}$ is the composition of the derivatives of the functions φ and ψ at $\mathbf{p}$ and $\varphi(\mathbf{p})$ respectively.

Proof

The differentiability of the functions φ and ψ at $\mathbf{p}$ and $\varphi(\mathbf{p})$ respectively ensures that there exist positive real numbers M, N, δ_1 and η_1 such that the following conditions hold: $\mathbf{x} \in X$ and $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}|$ for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta_1$; $\mathbf{y} \in Y$ and $|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p}))| \leq N\,|\mathbf{y} - \varphi(\mathbf{p})|$ for all $\mathbf{y} \in \mathbb{R}^n$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_1$; $|(D\psi)_{\varphi(\mathbf{p})}\mathbf{w}| \leq N\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^n$. (This follows on applying Proposition C.)

Let some positive real number ε be given. It follows from the differentiability of ψ at $\varphi(\mathbf{p})$ that there exists some real number η_2, where $0 < \eta_2 \leq \eta_1$, such that

$$\left|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\mathbf{y} - \varphi(\mathbf{p}))\right| \leq \frac{\varepsilon}{2M}|\mathbf{y} - \varphi(\mathbf{p})|$$

for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_2$. (This follows from a direct application of Lemma B.)

Let some positive real number ε be given. It follows from the differentiability of ψ at $\varphi(\mathbf{p})$ that there exists some real number η_2, where $0 < \eta_2 \leq \eta_1$, such that

$$\left|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\mathbf{y} - \varphi(\mathbf{p}))\right| \leq \frac{\varepsilon}{2M}|\mathbf{y} - \varphi(\mathbf{p})|$$

for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_2$. (This follows from a direct application of Lemma B.) Let some real number δ_2 be chosen so that $0 < \delta_2 \leq \delta_1$ and $M\delta_2 \leq \eta_2$.

Let some positive real number ε be given. It follows from the differentiability of ψ at $\varphi(\mathbf{p})$ that there exists some real number η_2, where $0 < \eta_2 \leq \eta_1$, such that

$$\left|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\mathbf{y} - \varphi(\mathbf{p}))\right| \leq \frac{\varepsilon}{2M}|\mathbf{y} - \varphi(\mathbf{p})|$$

for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_2$. (This follows from a direct application of Lemma B.) Let some real number δ_2 be chosen so that $0 < \delta_2 \leq \delta_1$ and $M\delta_2 \leq \eta_2$. If $\mathbf{x} \in \mathbb{R}^m$ satisfies $|\mathbf{x} - \mathbf{p}| < \delta_2$ then $\mathbf{x} \in X$ and $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}| < \eta_2$.

Let some positive real number ε be given. It follows from the differentiability of ψ at $\varphi(\mathbf{p})$ that there exists some real number η_2, where $0 < \eta_2 \leq \eta_1$, such that

$$\left|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\mathbf{y} - \varphi(\mathbf{p}))\right| \leq \frac{\varepsilon}{2M}|\mathbf{y} - \varphi(\mathbf{p})|$$

for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_2$. (This follows from a direct application of Lemma B.) Let some real number δ_2 be chosen so that $0 < \delta_2 \leq \delta_1$ and $M\delta_2 \leq \eta_2$. If $\mathbf{x} \in \mathbb{R}^m$ satisfies $|\mathbf{x} - \mathbf{p}| < \delta_2$ then $\mathbf{x} \in X$ and $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}| < \eta_2$. Consequently if $|\mathbf{x} - \mathbf{p}| < \delta$ then

$$\begin{aligned}
&\left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))\right| \\
&\qquad \leq \frac{\varepsilon}{2M}|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \\
&\qquad \leq \tfrac{1}{2}\,\varepsilon|\mathbf{x} - \mathbf{p}|.
\end{aligned}$$

Now it follows from the differentiability of φ at $\mathbf{p}$ that there exists some real number δ satisfying the inequalities $0 < \delta \leq \delta_2$ that is small enough to ensure that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \frac{\varepsilon}{2N}|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

Now it follows from the differentiability of φ at $\mathbf{p}$ that there exists some real number δ satisfying the inequalities $0 < \delta \leq \delta_2$ that is small enough to ensure that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \frac{\varepsilon}{2N}|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. Now $|(D\psi)_{\varphi(\mathbf{p})}\mathbf{w}| \leq N\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^n$.

Now it follows from the differentiability of φ at $\mathbf{p}$ that there exists some real number δ satisfying the inequalities $0 < \delta \leq \delta_2$ that is small enough to ensure that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \frac{\varepsilon}{2N}|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. Now $|(D\psi)_{\varphi(\mathbf{p})}\mathbf{w}| \leq N\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^n$. It follows that

$$\begin{aligned} \left|(D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\right| \\ &\leq N\,|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \\ &\leq \tfrac{1}{2}\varepsilon|\mathbf{x} - \mathbf{p}| \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

The inequalities obtained above ensure that $\mathbf{x} \in X$ and

$$
\begin{aligned}
&\left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\right| \\
&\qquad \leq \quad \left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))\right| \\
&\qquad\qquad + \left|(D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\right| \\
&\qquad \leq \quad \varepsilon|\mathbf{x} - \mathbf{p}|
\end{aligned}
$$

at all points $\mathbf{x}$ of $\mathbb{R}^m$ that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$.

The inequalities obtained above ensure that $\mathbf{x} \in X$ and

$$\begin{aligned} &\left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\right| \\ &\qquad \leq \left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))\right| \\ &\qquad\qquad + \left|(D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\right| \\ &\qquad \leq \varepsilon|\mathbf{x} - \mathbf{p}| \end{aligned}$$

at all points $\mathbf{x}$ of $\mathbb{R}^m$ that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. It follows from this that the composition function $\psi \circ \varphi$ is differentiable at $\mathbf{p}$, and that $(D(\psi \circ \varphi))_{\mathbf{p}} = (D\psi)_{\varphi(\mathbf{p})} \circ (D\varphi)_{\mathbf{p}}$, as required. ∎

An Alternative Proof of the Chain Rule

Let $\mathbf{q} = \varphi(\mathbf{p})$, and let $\sigma\colon X \to \mathbb{R}^n$ and $\tau\colon Y \to \mathbb{R}^k$ be the uniquely-determined functions defined throughout the domains X and Y of the functions φ and ψ respectively so that $\sigma(\mathbf{p}) = \mathbf{0}$, $\tau(\mathbf{q}) = \mathbf{0}$,

$$\varphi(\mathbf{x}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + |\mathbf{x} - \mathbf{p}|\,\sigma(\mathbf{x})$$

for all points $\mathbf{x}$ of the domain X of the function φ, and

$$\psi(\mathbf{y}) = \psi(\mathbf{q}) + (D\psi)_{\mathbf{q}}(\mathbf{y} - \mathbf{q}) + |\mathbf{y} - \mathbf{q}|\,\tau(\mathbf{y})$$

for all points $\mathbf{y}$ of the domain Y of the function ψ.

An Alternative Proof of the Chain Rule

Let $\mathbf{q} = \varphi(\mathbf{p})$, and let $\sigma\colon X \to \mathbb{R}^n$ and $\tau\colon Y \to \mathbb{R}^k$ be the uniquely-determined functions defined throughout the domains X and Y of the functions φ and ψ respectively so that $\sigma(\mathbf{p}) = \mathbf{0}$, $\tau(\mathbf{q}) = \mathbf{0}$,

$$\varphi(\mathbf{x}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + |\mathbf{x} - \mathbf{p}|\, \sigma(\mathbf{x})$$

for all points $\mathbf{x}$ of the domain X of the function φ, and

$$\psi(\mathbf{y}) = \psi(\mathbf{q}) + (D\psi)_{\mathbf{q}}(\mathbf{y} - \mathbf{q}) + |\mathbf{y} - \mathbf{q}|\, \tau(\mathbf{y})$$

for all points $\mathbf{y}$ of the domain Y of the function ψ. The differentiability of the functions φ and ψ at the points $\mathbf{p}$ and $\mathbf{q}$ then ensures that the functions σ and τ are continuous at the points $\mathbf{p}$ and $\mathbf{q}$ respectively, where $\mathbf{q} = \varphi(\mathbf{p})$ (see Lemma A).

An Alternative Proof of the Chain Rule

Let $\mathbf{q} = \varphi(\mathbf{p})$, and let $\sigma\colon X \to \mathbb{R}^n$ and $\tau\colon Y \to \mathbb{R}^k$ be the uniquely-determined functions defined throughout the domains X and Y of the functions φ and ψ respectively so that $\sigma(\mathbf{p}) = \mathbf{0}$, $\tau(\mathbf{q}) = \mathbf{0}$,

$$\varphi(\mathbf{x}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + |\mathbf{x} - \mathbf{p}|\,\sigma(\mathbf{x})$$

for all points $\mathbf{x}$ of the domain X of the function φ, and

$$\psi(\mathbf{y}) = \psi(\mathbf{q}) + (D\psi)_{\mathbf{q}}(\mathbf{y} - \mathbf{q}) + |\mathbf{y} - \mathbf{q}|\,\tau(\mathbf{y})$$

for all points $\mathbf{y}$ of the domain Y of the function ψ. The differentiability of the functions φ and ψ at the points $\mathbf{p}$ and $\mathbf{q}$ then ensures that the functions σ and τ are continuous at the points $\mathbf{p}$ and $\mathbf{q}$ respectively, where $\mathbf{q} = \varphi(\mathbf{p})$ (see Lemma A). Moreover the composition function $\tau \circ \varphi$ is continuous at the point $\mathbf{p}$, because the functions φ and τ are continuous at the points $\mathbf{p}$ and $\varphi(\mathbf{p})$ respectively.

The linearity of $(D\psi)_{\mathbf{q}} \colon \mathbb{R}^n \to \mathbb{R}^k$ then ensures that

$$\begin{aligned}
\psi(\varphi(\mathbf{x})) &= \psi(\mathbf{q}) + (D\psi)_{\mathbf{q}}(\varphi(\mathbf{x}) - \mathbf{q})) + |\varphi(\mathbf{x}) - \mathbf{q}|\,\tau(\varphi(\mathbf{x})) \\
&= \psi(\varphi(\mathbf{p})) + (D\psi)_{\mathbf{q}}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))) \\
&\qquad + |\varphi(\mathbf{x}) - \varphi(\mathbf{p})|\,\tau(\varphi(\mathbf{x})) \\
&= \psi(\varphi(\mathbf{p})) + (D\psi)_{\mathbf{q}}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) \\
&\qquad + |\mathbf{x} - \mathbf{p}|(D\psi)_{\mathbf{q}}(\sigma(\mathbf{x})) + |\varphi(\mathbf{x}) - \varphi(\mathbf{p})|\,\tau(\varphi(\mathbf{x})) \\
&= \psi(\varphi(\mathbf{p})) + (D\psi)_{\mathbf{q}}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + |\mathbf{x} - \mathbf{p}|\chi(\mathbf{x})
\end{aligned}$$

for all $\mathbf{x} \in X$, where $\chi \colon X \to \mathbb{R}^k$ is the uniquely-determined function on the domain X of the function φ defined so that $\chi(\mathbf{p}) = 0$ and

$$\chi(\mathbf{x}) = (D\psi)_{\mathbf{q}}(\sigma(\mathbf{x})) + \frac{|\varphi(\mathbf{x}) - \varphi(\mathbf{p})|}{|\mathbf{x} - \mathbf{p}|}\,\tau(\varphi(\mathbf{x}))$$

for all points $\mathbf{x}$ of the set X that are distinct from the point $\mathbf{p}$.

Thus, in order to complete the proof of the differentiability of the composition function $\psi \circ \varphi$ at the point $\mathbf{p}$, it suffices to show that that the function χ is continuous at the point $\mathbf{p}$ (see Lemma A), and moreover the continuity of the function χ at the point $\mathbf{p}$ can be established by verifying that $\lim_{\mathbf{x} \to \mathbf{p}} \chi(\mathbf{p}) = \mathbf{0}$.

Now $\lim_{\mathbf{x}\to\mathbf{p}} \sigma(\mathbf{x}) = \mathbf{0}$. The continuity of the linear transformation $(D\psi)_{\mathbf{q}}$ therefore ensures that

$$\lim_{\mathbf{x}\to\mathbf{p}} (D\psi)_{\mathbf{q}}(\sigma(\mathbf{x})) = (D\psi)_{\mathbf{q}} \left(\lim_{\mathbf{x}\to\mathbf{p}} \sigma(\mathbf{x}) \right) = (D\psi)_{\mathbf{q}}(\mathbf{0}) = \mathbf{0}.$$

Also there exist positive real numbers M and δ_0 such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}|$ whenever $|\mathbf{x} - \mathbf{p}| < \delta_0$ (see Proposition C). Then, given any positive real number ε, there exists some real number δ satisfying $0 < \delta < \delta_0$ which is small enough to ensure that $|\tau(\varphi(\mathbf{x}))| < \varepsilon/M$ whenever $|\mathbf{x} - \mathbf{p}| < \delta$, because $\tau(\varphi(\mathbf{p})) = \tau(\mathbf{q}) = \mathbf{0}$ and the composition function $\tau \circ \varphi$ is continuous at the point $\mathbf{p}$.

Also there exist positive real numbers M and δ_0 such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}|$ whenever $|\mathbf{x} - \mathbf{p}| < \delta_0$ (see Proposition C). Then, given any positive real number ε, there exists some real number δ satisfying $0 < \delta < \delta_0$ which is small enough to ensure that $|\tau(\varphi(\mathbf{x}))| < \varepsilon/M$ whenever $|\mathbf{x} - \mathbf{p}| < \delta$, because $\tau(\varphi(\mathbf{p})) = \tau(\mathbf{q}) = \mathbf{0}$ and the composition function $\tau \circ \varphi$ is continuous at the point $\mathbf{p}$. It follows that

$$\frac{|\varphi(\mathbf{x}) - \varphi(\mathbf{p})|}{|\mathbf{x} - \mathbf{p}|} |\tau(\varphi(\mathbf{x}))| < \varepsilon$$

whenever $|\mathbf{x} - \mathbf{p}| < \delta$. Consequently

$$\lim_{\mathbf{x} \to \mathbf{p}} \left(\frac{|\varphi(\mathbf{x}) - \varphi(\mathbf{p})|}{|\mathbf{x} - \mathbf{p}|} |\tau(\varphi(\mathbf{x}))| \right) = \mathbf{0}.$$

We can now conclude that

$$\begin{aligned}\lim_{\mathbf{x}\to\mathbf{p}} \chi(\mathbf{x}) &= \lim_{\mathbf{x}\to\mathbf{p}} (D\psi)_{\mathbf{q}}(\sigma(\mathbf{x})) + \lim_{\mathbf{x}\to\mathbf{p}} \left(\frac{|\varphi(\mathbf{x}) - \varphi(\mathbf{p})|}{|\mathbf{x}-\mathbf{p}|} \, \tau(\varphi(\mathbf{x})) \right) \\ &= \mathbf{0} = \chi(\mathbf{p}),\end{aligned}$$

and consequently the composition function $\psi \circ \varphi$ is differentiable at the point $\mathbf{p}$, with derivative $(D\psi)_{\varphi(\mathbf{p})} \circ (D\varphi)_{\mathbf{p}}$, as required. ∎