2 Differentiation of Functions of Several Variables

2021-22

2.1 Directional Derivatives

Recall the definition that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if, and only if,

$$\lim_{t\to 0} \frac{f(a+t) - f(a)}{t}$$

exists. This cannot be naively extended to vector-valued functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ of severable variables by looking at

$$\lim_{\mathbf{t}\to \mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{t})}{\mathbf{t}}$$

since we cannot divide by a vector $\mathbf{t} \in \mathbb{R}^n$. However, just as we looked at limits along straight lines, we can first restrict to $\mathbf{t} \in \mathbb{R}^n$ lying on a straight line through the origin, i.e. $\mathbf{t} = t\mathbf{v}$, and define a *directional derivative.*

Definition 1 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is a vector-valued function defined on an open set* U *and* $\mathbf{a} \in U$*. Given a unit vector* $\mathbf{v} \in \mathbb{R}^n$ *then the* ***directional derivative of*** $\mathbf{f}$ ***at*** $\mathbf{a}$ ***in the direction*** $\mathbf{v}$ *is defined to be*

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = \lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{v}) - \mathbf{f}(\mathbf{a})}{t},$$

an element of $\mathbb{R}^m$*, if it exists.*

If $d_{\mathbf{v}}\mathbf{f}(\mathbf{a})$ exists for all $\mathbf{a} \in U$ we have a *function* $d_{\mathbf{v}}\mathbf{f} : U \to \mathbb{R}^m$.

Note i many authors do **not** require that $\mathbf{v}$ be a *unit* vector, and the definition would still be meaningful for a non-unit vector. I like the *directional* derivative to depend **only** on the direction of $\mathbf{v}$ which why I assume that $\mathbf{v}$ is a unit vector.

Note ii The directional **limit** had $t \to 0+$, a one-sided limit; the directional **derivative** has $t \to 0$. For directional limits I was interested to know if the directional limit along $\mathbf{v}$ was different from that along $-\mathbf{v}$; if they were different it would then imply the limit did not exist. For directional derivatives

I want to know that both one-sided limits at 0 exist and are equal, i.e. the limit is as $t \to 0$.

If f is a scalar-valued function the definition is easy to verify:

Example 2 *Let* $f(\mathbf{x}) = xy^2$ *for* $\mathbf{x} = (x,y)^T \in \mathbb{R}^2$*. Find, from first principles, the directional derivative of* f *at* $\mathbf{a} = (2,1)^T$ *in the direction of the unit vector* $\mathbf{v} = (1,-1)^T/\sqrt{2}$*.*

Solution Consider

$$\begin{aligned}
\frac{f(\mathbf{a}+t\mathbf{v}) - f(\mathbf{a})}{t} &= \frac{1}{t}\left(f\left(\begin{pmatrix}2\\1\end{pmatrix} + \frac{t}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\right) - f\left(\begin{pmatrix}2\\1\end{pmatrix}\right)\right) \\
&= \frac{1}{t}\left(f\left(\begin{pmatrix}2+t/\sqrt{2}\\1-t/\sqrt{2}\end{pmatrix}\right) - f\left(\begin{pmatrix}2\\1\end{pmatrix}\right)\right) \\
&= \frac{1}{t}\left(\left(2+\frac{t}{\sqrt{2}}\right)\left(1-\frac{t}{\sqrt{2}}\right)^2 - 2\right) \\
&= \frac{1}{t}\left(\left(\frac{1}{\sqrt{2}} - \frac{4}{\sqrt{2}}\right)t + \frac{t^3}{2\sqrt{2}}\right) \\
&\to -\frac{3}{\sqrt{2}}
\end{aligned}$$

as $t \to 0$. Hence

$$d_{\mathbf{v}}f(\mathbf{a}) = -\frac{3}{\sqrt{2}}.$$

■

The following result reduces directional derivatives of *vector*-valued functions to directional derivatives of *scalar*-valued functions.

Proposition 3 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is a function defined on an open set* U *with component functions* $f^i : U \subseteq \mathbb{R}^n \to \mathbb{R}$, $1 \le i \le m$*. Then, for unit* $\mathbf{v} \in \mathbb{R}^n$ *and* $\mathbf{a} \in U$,

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) \text{ exists } \iff \forall\, 1 \le i \le m,\ d_{\mathbf{v}}f^i(\mathbf{a}) \text{ exists.}$$

Further

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) \text{ exists } \implies \forall\, 1 \le i \le m,\ \left(d_{\mathbf{v}}\mathbf{f}(\mathbf{a})\right)^i = d_{\mathbf{v}}f^i(\mathbf{a}).$$

Proof Follows from the result in Section 1 that $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x}) = \mathbf{b}$ if, and only if, $\lim_{\mathbf{x}\to\mathbf{a}} g^i(\mathbf{x}) = b^i$ for all components $1 \le i \le m$. ■

Example 4 *Let* $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} xy^2 \\ x^2y \end{pmatrix},$$

for $\mathbf{x} = (x, y)^T \in \mathbb{R}^2$. *Find, from first principles, the directional derivative of* $\mathbf{f}$ *at* $\mathbf{a} = (2, 1)^T$ *in the direction of the unit vector* $\mathbf{v} = (1, -1)^T/\sqrt{2}$.

Solution *Question Sheet.* But, from Proposition 3 it suffices to find the directional derivative of each component function. The first component function was the subject of Example 2. So we are leaving to the student the calculation of the directional derivative of $(x, y)^T \mapsto x^2y$. ■

Proposition 5 *If* $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ *is a linear map and* $\mathbf{v} \in \mathbb{R}^n$ *is a unit vector then* $d_{\mathbf{v}}\mathbf{L}$ *exists and* $d_{\mathbf{v}}\mathbf{L}(\mathbf{a}) = \mathbf{L}(\mathbf{v})$ *for all* $\mathbf{a} \in \mathbb{R}^n$.

Proof Let $\mathbf{a}$ and unit $\mathbf{v} \in \mathbb{R}^n$ be given. Then for all $t \neq 0$ we have

$$\frac{\mathbf{L}(\mathbf{a} + t\mathbf{v}) - \mathbf{L}(\mathbf{a})}{t} = \frac{\mathbf{L}(\mathbf{a}) + t\mathbf{L}(\mathbf{v}) - \mathbf{L}(\mathbf{a})}{t} = \mathbf{L}(\mathbf{v}).$$

Hence $d_{\mathbf{v}}\mathbf{L}(\mathbf{a}) = \mathbf{L}(\mathbf{v})$. ■

2.2 Partial Derivatives

Recall that the **Standard basis for** $\mathbb{R}^n$ is $\{\mathbf{e}_i\}_{1\le i\le n}$ where $\mathbf{e}_i$ has 0 in all coordinates except the i-th where it has a 1. The directional derivatives in the direction of the basis vectors are the well known *partial derivatives*:

Definition 6 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is a function defined on an open set* U *and* $\mathbf{a} \in U$. *For* $1 \le j \le n$, *the* j***-th partial derivative of*** $\mathbf{f}$ ***at*** $\mathbf{a}$ *is defined to be the directional derivative in the direction of the standard basis vector* $\mathbf{e}_j$:

$$\frac{\partial \mathbf{f}}{\partial x^j}(\mathbf{a}) = d_{\mathbf{e}_j}\mathbf{f}(\mathbf{a}),$$

an element of $\mathbb{R}^m$, *if it exists.*

We also write $d_j\mathbf{f}(\mathbf{a})$ as shorthand for $d_{\mathbf{e}_j}\mathbf{f}(\mathbf{a})$, and thus $\partial\mathbf{f}(\mathbf{a})/\partial x^j$. So, to stress the point, the following are interchangeable

$$\boxed{d_j\mathbf{f}(\mathbf{a}) = d_{\mathbf{e}_j}\mathbf{f}(\mathbf{a}) = \frac{\partial\mathbf{f}}{\partial x^j}(\mathbf{a})\,.}$$

Assume that for $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{a} \in U$ the directional derivative $d_j\mathbf{f}(\mathbf{a})$ exists. Then Proposition 3 says that for $1 \le i \le m$ the i-th component of $d_j\mathbf{f}(\mathbf{a})$ equals the directional derivative of the i-th component of f, i.e. $(d_j\mathbf{f}(\mathbf{a}))^i = d_jf^i(\mathbf{a})$. These can be put together as

$$d_j\mathbf{f}(\mathbf{a}) = \begin{pmatrix} d_jf^1(\mathbf{a}) \\ d_jf^2(\mathbf{a}) \\ \vdots \\ d_jf^m(\mathbf{a}) \end{pmatrix}, \quad \text{that is} \quad \frac{\partial\mathbf{f}}{\partial x^j}(\mathbf{a}) = \begin{pmatrix} \partial f^1(\mathbf{a})/\partial x^j \\ \partial f^2(\mathbf{a})/\partial x^j \\ \vdots \\ \partial f^m(\mathbf{a})/\partial x^j \end{pmatrix}.$$

Note that it is possible that the partial derivatives $d_j\mathbf{f}(\mathbf{a})$ exist for all $1 \le j \le n$ yet $d_\mathbf{v}\mathbf{f}(\mathbf{a})$ does not exist for some unit vector $\mathbf{v}$. That is

$$\boxed{\forall\, 1 \le j \le n,\ d_j\mathbf{f}(\mathbf{a}) \text{ exists } \not\Longrightarrow \forall\, \text{unit } \mathbf{v},\ d_\mathbf{v}\mathbf{f}(\mathbf{a}) \text{ exists.}}$$

See the Question Sheet for an example.

2.3 Fréchet Derivative

As motivation for what follows, given a differentiable function $f : \mathbb{R} \to \mathbb{R}$ rearrange the definition of derivative

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

as

$$\lim_{t\to 0} \frac{f(a+t) - f(a) - f'(a)\,t}{t} = 0. \tag{1}$$

Here, for fixed $a \in \mathbb{R}$, $t \mapsto f'(a)\,t$ is a linear function $\mathbb{R} \to \mathbb{R}$, which could be denoted as $L_a(t) = f'(a)\,t$.

Note that, because the limit is 0, we have

$$\lim_{t\to 0}\frac{g(t)}{t}=0\Longleftrightarrow\lim_{t\to 0}\left|\frac{g(t)}{t}\right|=0\Longleftrightarrow\lim_{t\to 0}\left|\frac{g(t)}{|t|}\right|=0\Longleftrightarrow\lim_{t\to 0}\frac{g(t)}{|t|}=0.$$

So (1) holds if, and only if,

$$\lim_{t\to 0}\frac{f(a+t)-f(a)-f'(a)\,t}{|t|}=0. \tag{2}$$

It is the characterisation of the derivative as the linear function $L_a(t)$ satisfying

$$\lim_{t\to 0}\frac{f(a+t)-f(a)-L_a(t)}{|t|}=0.$$

which can be generalised.

Definition 7 *The vector-valued function* $\mathbf{f}:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$ *is* ***Fréchet differentiable at*** $\mathbf{a}\in U$ *if, and only if, there exists a linear map* $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ *such that*

$$\lim_{\mathbf{t}\to\mathbf{0}}\frac{\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a})-\mathbf{L}(\mathbf{t})}{|\mathbf{t}|}=\mathbf{0}.$$

The linear function $\mathbf{L}$ *is called the* ***Fréchet derivative of*** $\mathbf{f}$ ***at*** $\mathbf{a}$ *and is denoted by* $d\mathbf{f}_{\mathbf{a}}$.

Definition 8 *If* $\mathbf{f}$ *is differentiable at each point of* U *we say that* $\mathbf{f}$ *is a* ***Fréchet differentiable function on*** U.

The following result is **important** but, due to a lack of time, I do not give a proof in lectures.

Theorem 9 *If* $\mathbf{f}:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$, *a function defined on an open set* U, *is Fréchet differentiable at* $\mathbf{a}\in U$ *then the derivative is unique.*

Proof Not given but see the appendix. ■

Verifying that the definition holds for a *vector*-valued function is time-consuming. The following result reduces the question of a vector-valued function being Fréchet differentiable to that of each of its *scalar*-valued component functions being Fréchet differentiable. Compare with Proposition 3.

Proposition 10 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is a function defined on an open set* U *with component functions* $f^i : U \to \mathbb{R}$, $1 \le i \le m$, *and* $\mathbf{a} \in U$. *Then*

$$d\mathbf{f}_{\mathbf{a}} \text{ exists} \iff \forall\, 1 \le i \le m,\ df^i_{\mathbf{a}} \text{ exists.}$$

Further

$$d\mathbf{f}_{\mathbf{a}} \text{ exists} \implies \forall\, 1 \le i \le m,\ \ (d\mathbf{f}_{\mathbf{a}})^i = df^i_{\mathbf{a}}.$$

Proof (*Not in lectures*) Recall from Chapter 1 that $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x}) = \mathbf{b}$ if, and only if, $\lim_{\mathbf{x}\to\mathbf{a}} g^i(\mathbf{x}) = b^i$ for all components $1 \le i \le m$. Then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{x}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{x})}{|\mathbf{x}|} = \mathbf{0},$$

if, and only if,

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{f^i(\mathbf{a}+\mathbf{x}) - f^i(\mathbf{a}) - (d\mathbf{f}_{\mathbf{a}})^i(\mathbf{x})}{|\mathbf{x}|} = 0,$$

for $1 \le i \le m$. That is, $\mathbf{f}$ is Fréchet differentiable at $\mathbf{a}$ if, and only if, for all $1 \le i \le m$ each f^i is Fréchet differentiable at $\mathbf{a}$.

But further, if $\mathbf{f}$ is Fréchet differentiable at $\mathbf{a}$ then $(d\mathbf{f}_{\mathbf{a}})^i$ satisfies the definition for the derivative of f^i. Yet, by Theorem 9, the Fréchet derivative of f^i is **unique** and is defined to be $df^i_{\mathbf{a}}$. Hence $(d\mathbf{f}_{\mathbf{a}})^i = df^i_{\mathbf{a}}$.

■

We can now give some examples for scalar-valued functions.

Example 11 *If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ (in the manner before this course) then f has a Fréchet derivative at a with $df_a(t) = f'(a)\,t$ for all $t \in \mathbb{R}$.*

Solution This is simply (2) above:

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

■

For a particular example, the function $f(x) = x^3$ is Fréchet differentiable for all $a \in \mathbb{R}$ with Fréchet derivative $df_a(t) = 3a^2 t$ for all $t \in \mathbb{R}$.

Returning to Example 2, when we last looked at directional derivatives.

Example 2 continued *Let $f(\mathbf{x}) = xy^2$ on $\mathbb{R}^2$. By verifying the definition show that f is Fréchet differentiable at $\mathbf{a} = (2,-1)^T$ and find the Fréchet derivative, $df_{\mathbf{a}}(\mathbf{t})$.*

Solution With $\mathbf{t} = (s,t)^T$ we have

$$\begin{aligned} f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) &= f\left(\begin{pmatrix} 2+s \\ -1+t \end{pmatrix}\right) - f\left(\begin{pmatrix} 2 \\ -1 \end{pmatrix}\right) \\ &= (2+s)(-1+t)^2 - 2 \\ &= s - 4t + 2t^2 - 2st + st^2. \end{aligned}$$

The 'linear part' of this (i.e. the linear combination of the variables s and t) is $s - 4t$, so we *guess* $df_{\mathbf{a}}(\mathbf{t}) = s - 4t$.

To *check*,

$$\frac{f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - (s-4t)}{|\mathbf{t}|} = \frac{2t^2 - 2st + st^2}{|\mathbf{t}|}.$$

As noted before, $|s|, |t| \le |\mathbf{t}|$ so, by the triangle inequality,

$$\begin{aligned} \left|2t^2 - 2st + st^2\right| &\le 2|t|^2 + 2|s||t| + |s||t|^2 \\ &\le 4|\mathbf{t}|^2 + |\mathbf{t}|^3. \end{aligned}$$

Therefore

$$\left|\frac{2t^2-2st+st^2}{|\mathbf{t}|}\right| \leq 4\left|\mathbf{t}\right|+\left|\mathbf{t}\right|^2 \to 0$$

as $\mathbf{t}\to\mathbf{0}$. Thus, by the Sandwich Rule,

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{2t^2-2st+st^2}{|\mathbf{t}|}=0.$$

Hence f is Fréchet differentiable on $\mathbb{R}^2$ and $df_{\mathbf{a}}(\mathbf{t}) = s-4t$ for $\mathbf{t}=(s,t)^T$. ■

We can repeat this, but for a general point $\mathbf{a}$:

Example 2 continued *Let $f(\mathbf{x}) = xy^2$ on $\mathbb{R}^2$. By verifying the definition show that f is Fréchet differentiable at $\mathbf{a}\in\mathbb{R}^2$ and the Fréchet derivative is*

$$df_{\mathbf{a}}(\mathbf{t}) = \beta^2 s + 2\alpha\beta t,$$

if $\mathbf{a}=(\alpha,\beta)^T$ and $\mathbf{t}=(s,t)^T$.

Solution *Problems Class* Writing $\mathbf{a}=(\alpha,\beta)^T$ and $\mathbf{t}=(s,t)^T$ think of the upper case α and β as fixed while the lower case s,t vary (are the variables). We have

$$\begin{aligned} f(\mathbf{a}+\mathbf{t})-f(\mathbf{a}) &= f\left(\begin{pmatrix}\alpha+s\\ \beta+t\end{pmatrix}\right)-f\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)\\ &= \beta^2 s+2\alpha\beta t+\alpha t^2+2\beta st+st^2. \end{aligned}$$

The 'linear part' of function of the variables s and t is $\beta^2 s+2\alpha\beta t$, so we *guess*

$$df_{\mathbf{a}}(\mathbf{t}) = \beta^2 s+2\alpha\beta t.$$

To *check*,

$$\frac{f(\mathbf{a}+\mathbf{t})-f(\mathbf{a})-\left(\beta^2 s+2\alpha\beta t\right)}{|\mathbf{t}|} = \frac{\alpha t^2+2\beta st+st^2}{|\mathbf{t}|}.$$

As noted before, $|s|,|t|\leq|\mathbf{t}|$ and $|\alpha|,|\beta|\leq|\mathbf{a}|$, so by the triangle inequality

$$\begin{aligned} \left|\alpha t^2+2\beta st+st^2\right| &\leq |\alpha|\left|t\right|^2+2\left|B\right|\left|s\right|\left|t\right|+\left|s\right|\left|t\right|^2\\ &\leq 3\left|\mathbf{a}\right|\left|\mathbf{t}\right|^2+\left|\mathbf{t}\right|^3. \end{aligned}$$

Therefore

$$\left|\frac{\alpha t^2 + 2\beta st + st^2}{|\mathbf{t}|}\right| \le 3\,|\mathbf{a}|\,|\mathbf{t}| + |\mathbf{t}|^2 \to 0$$

as $\mathbf{t} \to \mathbf{0}$. Thus, by the Sandwich Rule,

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{\alpha t^2 + 2\beta st + st^2}{|\mathbf{t}|} = 0.$$

Hence f is differentiable on $\mathbb{R}^2$ and

$$df_{\mathbf{a}}(\mathbf{t}) = \beta^2 s + 2\alpha\beta t,$$

where $\mathbf{a} = (\alpha, \beta)^T$ and $\mathbf{t} = (s, t)^T$. ■

In Example 4 the function xy^2 in this example can be just one of the coordinate functions of a vector valued function as in

Example 4 **continued** *Let* $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} xy^2 \\ x^2 y \end{pmatrix},$$

for $\mathbf{x} \in \mathbb{R}^2$. *Show that* $\mathbf{f}$ *is Fréchet differentiable on* $\mathbb{R}^2$ *and find the derivative,* $d\mathbf{f}_{\mathbf{a}}(\mathbf{t})$ *for* $\mathbf{a} \in \mathbb{R}^2$.

Solution *Question Sheet,* but by Proposition 10 it suffices to show that each scalar-valued component function is Fréchet differentiable on $\mathbb{R}^2$ and find their Fréchet derivatives.

For $f^1(\mathbf{x}) = xy^2$ this was done in Example 2, so if $\mathbf{a} = (\alpha, \beta)^T$ and $\mathbf{t} = (s, t)^T$ then $df^1_{\mathbf{a}}(\mathbf{t}) = \beta^2 s + 2\alpha\beta t$.

For $f^2(\mathbf{x}) = x^2 y$ this is a question on the Problem Sheet with answer $df^2_{\mathbf{a}}(\mathbf{t}) = 2\alpha\beta s + \alpha^2 t$.

Since both f^1 and f^2 are Fréchet differentiable at $\mathbf{a}$ then so is $\mathbf{f}$, with Fréchet derivative

$$d\mathbf{f}_{\mathbf{a}}(\mathbf{t}) = \begin{pmatrix} df^1_{\mathbf{a}}(\mathbf{t}) \\ df^2_{\mathbf{a}}(\mathbf{t}) \end{pmatrix} = \begin{pmatrix} \beta^2 s + 2\alpha\beta t \\ 2\alpha\beta s + \alpha^2 t \end{pmatrix}.$$

■

Example 12 *For any* $\mathbf{k} \in \mathbb{R}^m$*, the* ***constant function*** $c_{\mathbf{k}} : \mathbb{R}^n \to \mathbb{R}^m$*,* $\mathbf{x} \mapsto \mathbf{k}$ *is Fréchet differentiable and* $d\left(c_{\mathbf{k}}\right)_{\mathbf{a}} = \mathbf{0}$ *for all* $\mathbf{a} \in \mathbb{R}^n$*.*

Solution $c_{\mathbf{k}}\left(\mathbf{a}+\mathbf{t}\right) - c_{\mathbf{k}}\left(\mathbf{a}\right) = \mathbf{k} - \mathbf{k} = \mathbf{0}$ and so $L\left(\mathbf{t}\right) = \mathbf{0}$. ■

The following is straightforward and will be used later. It says that *all linear functions are Fréchet differentiable.* Thus the Fréchet derivative exists and, being a linear function, there is an obvious candidate for what it should be, namely the linear function you started with!

Example 13 *All linear functions* $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ *are Fréchet differentiable on* $\mathbb{R}^n$ *and* $d\mathbf{L}_{\mathbf{a}} = \mathbf{L}$ *for all* $\mathbf{a} \in \mathbb{R}^n$*.*

Solution Since $\mathbf{L}$ is linear, $\mathbf{L}(\mathbf{a}+\mathbf{x}) - \mathbf{L}(\mathbf{a}) = \mathbf{L}(\mathbf{x})$. Hence, for $\mathbf{x} \neq \mathbf{0}$,

$$\frac{\mathbf{L}(\mathbf{a}+\mathbf{x}) - \mathbf{L}(\mathbf{a}) - d\mathbf{L}_{\mathbf{a}}(\mathbf{x})}{|\mathbf{x}|} = \frac{\mathbf{L}(\mathbf{x}) - d\mathbf{L}_{\mathbf{a}}(\mathbf{x})}{|\mathbf{x}|}.$$

With the choice of $d\mathbf{L}_{\mathbf{a}} = \mathbf{L}$ this is zero, and so the limit is zero. ■

2.4 Fréchet Differentiable implies Continuous

We have seen the following result in earlier courses on functions $f : A \subseteq \mathbb{R} \to \mathbb{R}$; if f is differentiable at a then it is continuous at a. The proof is almost identical but I give it since it does require the fact that linear functions are continuous at the origin.

Proposition 14 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, *where* U *is an open subset, is Fréchet differentiable at* $\mathbf{a} \in U$, *then* $\mathbf{f}$ *is continuous at* $\mathbf{a}$.

Proof Consider

$$\begin{aligned}\lim_{\mathbf{t}\to\mathbf{0}} \big(\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{t})\big) &= \lim_{\mathbf{t}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{t})}{|\mathbf{t}|} |\mathbf{t}| \\ &= \lim_{\mathbf{t}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{t})}{|\mathbf{t}|} \lim_{\mathbf{t}\to\mathbf{0}} |\mathbf{t}|,\end{aligned}$$

by the Product Rule for limits, allowable since both limits exist. Then

$$\lim_{\mathbf{t}\to\mathbf{0}} \big(\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{t})\big) = \mathbf{0} \times 0 = \mathbf{0}.$$

Yet, from Section 1 we have that the linear function $d\mathbf{f}_{\mathbf{a}}$ is everywhere continuous so $\lim_{\mathbf{t}\to\mathbf{0}} d\mathbf{f}_{\mathbf{a}}(\mathbf{t}) = d\mathbf{f}_{\mathbf{a}}(\mathbf{0}) = \mathbf{0}$. Hence

$$\lim_{\mathbf{t}\to\mathbf{0}} \big(\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a})\big) = \mathbf{0},$$

i.e. $\lim_{\mathbf{t}\to\mathbf{0}} \mathbf{f}(\mathbf{a}+\mathbf{t}) = \mathbf{f}(\mathbf{a})$, so $\mathbf{f}$ is continuous at $\mathbf{a}$. ■

The following will have been seen in any course on functions $f : \mathbb{R} \to \mathbb{R}$ but is also true in our more general case and can never be said too often.

f Fréchet differentiable at **a** $\Longrightarrow$ **f** continuous at **a**

f continuous at **a** $\not\Longrightarrow$ **f** Fréchet differentiable at **a**

2.5 Fréchet derivative exists implies directional derivative exists

The next result is important; it says that if a function is Fréchet differentiable at a point $\mathbf{a}$ then all the directional derivatives exist at that point. Further the directional derivatives can be calculated by $d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = d\mathbf{f}_{\mathbf{a}}(\mathbf{v})$ for all unit $\mathbf{v}$.

Proposition 15 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$*, where* U *is an open subset, is Fréchet differentiable at* $\mathbf{a} \in U$*. Then for all unit vectors* $\mathbf{v}$*, the directional derivatives* $d_{\mathbf{v}}\mathbf{f}(\mathbf{a})$ *exist and*

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = d\mathbf{f}_{\mathbf{a}}(\mathbf{v})\,.$$

Proof From the assumption that $\mathbf{f}$ is Fréchet differentiable at $\mathbf{a}$ we have that there exists a linear function $d\mathbf{f}_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{a}+\mathbf{t}) - \mathbf{f}(\mathbf{a}) - d\mathbf{f}_{\mathbf{a}}(\mathbf{t})}{|\mathbf{t}|} = \mathbf{0}.$$

Then with $\mathbf{t} = t\mathbf{v}$, where $\mathbf{v}$ is a unit vector, (so $|\mathbf{t}| = |t|\,|\mathbf{v}| = |t|$), and the fact that $d\mathbf{f}_{\mathbf{a}}$ is a *linear* function, we get

$$\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{v}) - \mathbf{f}(\mathbf{a}) - t d\mathbf{f}_{\mathbf{a}}(\mathbf{v})}{|t|} = \mathbf{0}.$$

As seen in the derivation of (2) above, because the limit is $\mathbf{0}$, this is equivalent to

$$\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{v}) - \mathbf{f}(\mathbf{a}) - t d\mathbf{f}_{\mathbf{a}}(\mathbf{v})}{t} = \mathbf{0},$$

i.e.

$$\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{v}) - \mathbf{f}(\mathbf{a})}{t} = d\mathbf{f}_{\mathbf{a}}(\mathbf{v})\,.$$

This first tells us that the limit exists, i.e. $d_{\mathbf{v}}\mathbf{f}(\mathbf{a})$ exists. But further, it says that $d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = d\mathbf{f}_{\mathbf{a}}(\mathbf{v})$. ■

Important Note.

$df_{\mathbf{a}}$ exists $\implies \forall$ unit $\mathbf{v} \in \mathbb{R}^n, d_{\mathbf{v}}f(\mathbf{a})$ exists
$\forall$ unit $\mathbf{v} \in \mathbb{R}^n$, $d_{\mathbf{v}}f(\mathbf{a})$ exists $\not\Longrightarrow$ $df_{\mathbf{a}}$ exists.

Given this, it follows that by restricting to $\mathbf{v} = \mathbf{e}_i$, the basis vectors, that

$$\boxed{\begin{array}{l} d\mathbf{f}_{\mathbf{a}} \text{ exists} \implies \forall i \ \dfrac{\partial \mathbf{f}}{\partial x^i}(\mathbf{a}) \text{ exists} \\ \forall i \ \dfrac{\partial \mathbf{f}}{\partial x^i}(\mathbf{a}) \text{ exists} \ \not\Longrightarrow d\mathbf{f}_{\mathbf{a}} \text{ exists.} \end{array}}$$

Before we look for conditions under which the converse of Proposition 15 holds we will examine the matrix associated with the Fréchet derivative, a linear map.

2.6 Jacobian Matrix

Recall, all linear functions $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ have a unique associated matrix $M \in M_{m,n}(\mathbb{R})$ such that $\mathbf{L}(\mathbf{t}) = M\mathbf{t}$ for all $\mathbf{t} \in \mathbb{R}^n$. In fact, the *columns* of M are $\mathbf{L}(\mathbf{e}_j), 1 \le j \le n$.

So, if $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is Fréchet differentiable at $\mathbf{a} \in U$ then there is a matrix associated with the Fréchet derivative $d\mathbf{f}_\mathbf{a}$, called $J\mathbf{f}(\mathbf{a})$. Thus

$$\boxed{\forall \mathbf{t} \in \mathbb{R}^n,\ d\mathbf{f}_\mathbf{a}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t}.}$$

That is,

$$\boxed{d\mathbf{f}_\mathbf{a} \text{ exists} \implies \forall \mathbf{t} \in \mathbb{R}^n,\ d\mathbf{f}_\mathbf{a}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t}}$$

The columns of $J\mathbf{f}(\mathbf{a})$ are $d\mathbf{f}_\mathbf{a}(\mathbf{e}_j)$. Yet, for a Fréchet differentiable function we have, from Proposition 15, that

$$d\mathbf{f}_\mathbf{a}(\mathbf{e}_j) = d_{\mathbf{e}_j}\mathbf{f}(\mathbf{a}) = d_j\mathbf{f}(\mathbf{a}).$$

Hence the j-th column of the matrix associated with $\mathbf{f}$ is the j-th partial derivative of $\mathbf{f}$ at $\mathbf{a}$.

This can be taken as a definition of a matrix for a function even when the function is **not** Fréchet differentiable!

Definition 16 *Assume that all the partial derivatives of* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *exist at* $\mathbf{a} \in U$. *The* ***Jacobian matrix of*** $\mathbf{f}$ ***at*** $\mathbf{a}$ *is given by*

$$J\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ d_1\mathbf{f}(\mathbf{a}) & d_2\mathbf{f}(\mathbf{a}) & \ldots & d_n\mathbf{f}(\mathbf{a}) \\ \downarrow & \downarrow & \downarrow & \downarrow \end{pmatrix} \in M_{m,n}(\mathbb{R}). \tag{3}$$

Definition 17 *If all partial derivatives of* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *exist at all points of* U *we can look upon the Jacobian matrix as a function* $J\mathbf{f} : U \to M_{m,n}$, $\mathbf{a} \mapsto J\mathbf{f}(\mathbf{a})$, *the* ***Jacobian function***.

Stress: As already noted, a function can have all partial derivatives at a point without having a Fréchet derivative there. So the class of functions with a Jacobian matrix **strictly** contains the class of all Fréchet differentiable functions.

Double stress: the most commonly seen error is that students incorrectly give the **transpose** of the Jacobian matrix.

Aside The matrix (3) can be expanded as hh

$$
\begin{aligned}
J\mathbf{f}(\mathbf{a}) &= \begin{pmatrix} d_1 f^1(\mathbf{a}) & d_2 f^1(\mathbf{a}) & d_3 f^1(\mathbf{a}) & \cdots & d_n f^1(\mathbf{a}) \\ d_1 f^2(\mathbf{a}) & d_2 f^2(\mathbf{a}) & d_3 f^2(\mathbf{a}) & & d_n f^2(\mathbf{a}) \\ d_1 f^3(\mathbf{a}) & d_2 f^3(\mathbf{a}) & & & \vdots \\ \vdots & & & & \vdots \\ d_1 f^m(\mathbf{a}) & d_2 f^m(\mathbf{a}) & \cdots & \cdots & d_n f^m(\mathbf{a}) \end{pmatrix} \\
&= \begin{pmatrix} {\partial f^1(\mathbf{a})}/{\partial x^1} & {\partial f^1(\mathbf{a})}/{\partial x^2} & {\partial f^1(\mathbf{a})}/{\partial x^3} & \ldots & {\partial f^1(\mathbf{a})}/{\partial x^n} \\ {\partial f^2(\mathbf{a})}/{\partial x^1} & {\partial f^2(\mathbf{a})}/{\partial x^2} & & & {\partial f^2(\mathbf{a})}/{\partial x^n} \\ {\partial f^3(\mathbf{a})}/{\partial x^1} & & & & \vdots \\ \vdots & & & & \vdots \\ {\partial f^m(\mathbf{a})}/{\partial x^1} & {\partial f^m(\mathbf{a})}/{\partial x^2} & \ldots & \ldots & {\partial f^m(\mathbf{a})}/{\partial x^n} \end{pmatrix} \qquad (4)
\end{aligned}
$$

Returning to an earlier example,

Example 4 *Let* $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} xy^2 \\ x^2 y \end{pmatrix}.$$

Then, since all partial derivatives exist,

$$J\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \partial(xy^2)/\partial x & \partial(xy^2)/\partial y \\ \partial(x^2y)/\partial x & \partial(x^2y)/\partial y \end{pmatrix} = \begin{pmatrix} y^2 & 2xy \\ 2xy & x^2 \end{pmatrix}, \qquad (5)$$

for all $\mathbf{x} \in \mathbb{R}^2$. ■

We have, in fact, already seen this before where, in Example 4, we were asked to show, by verifying the definition, that $\mathbf{f}$ is Fréchet differentiable on

$\mathbb{R}^2$ and to find it's Fréchet derivative. The result was that if $\mathbf{a} = (\alpha, \beta)^T \in \mathbb{R}^2$ then $d\mathbf{f}_\mathbf{a}$ exists and

$$d\mathbf{f}_\mathbf{a}(\mathbf{t}) = \begin{pmatrix} \beta^2 s + 2\alpha\beta t \\ 2\alpha\beta s + \alpha^2 t \end{pmatrix} = \begin{pmatrix} \beta^2 & 2\alpha\beta \\ 2\alpha\beta & \alpha^2 \end{pmatrix} \mathbf{t}, \tag{6}$$

where $\mathbf{t} = (s,t)^T$. But recall that if $d\mathbf{f}_\mathbf{a}$ exists then $d\mathbf{f}_\mathbf{a}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t}$. Hence (6) gives

$$J\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \beta^2 & 2\alpha\beta \\ 2\alpha\beta & \alpha^2 \end{pmatrix},$$

which matches (5) with $\mathbf{x} = \mathbf{a}$.

2.7 Special Case: functions of one variable

If $\mathbf{f} : U \subseteq \mathbb{R} \to \mathbb{R}^m$ is a function of **one** variable**,** t, then the Jacobian matrix is labelled $\mathbf{f}'(t)$ instead of $J\mathbf{f}(t)$. That is,

$$J\mathbf{f}(t) = \begin{pmatrix} df^1(t)/dt \\ \vdots \\ \vdots \\ df^m(t)/dt \end{pmatrix} = \mathbf{f}'(t).$$

say.

2.8 Special Case: Scalar-valued functions and the Gradient Vector

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is a **scalar-valued function** then $Jf(\mathbf{x})$ is a $1 \times n$ matrix and the transpose is an $n \times 1$ vector.

Definition 18 *Assume $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is a **scalar**-valued function defined on an open set U, all of whose partial derivatives exist at $\mathbf{a} \in U$. The **Gradient Vector of** f **at** $\mathbf{a}$ is given by*

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x^1}(\mathbf{a}), \ \frac{\partial f}{\partial x^2}(\mathbf{a}), \ \cdots, \ \frac{\partial f}{\partial x^n}(\mathbf{a}) \right)^T = Jf(\mathbf{a})^T.$$

*If the partial derivatives exist at all points of U then $\nabla f : U \to \mathbb{R}$, $\mathbf{a} \mapsto \nabla f(\mathbf{a})$ is the **gradient function of** f.*

Stress that the Gradient Vector is **only** defined for scalar-valued functions.

Then, for a Fréchet differentiable function f,

$$df_{\mathbf{a}}(\mathbf{t}) = Jf(\mathbf{a})\,\mathbf{t} = \nabla f(\mathbf{a})^T\,\mathbf{t} = \nabla f(\mathbf{a}) \bullet \mathbf{t}.$$

Example *Let* $f(\mathbf{x}) = xy^2$ *on* $\mathbb{R}^2$. *Then*

$$\nabla f(\mathbf{x}) = \begin{pmatrix} y^2 \\ 2xy \end{pmatrix}.$$

So, for $\mathbf{a} = (\alpha, \beta)^T$ and $\mathbf{t} = (s, t)^T$,

$$\nabla f(\mathbf{a}) \bullet \mathbf{t} = \begin{pmatrix} \beta^2 \\ 2\alpha\beta \end{pmatrix} \bullet \begin{pmatrix} s \\ t \end{pmatrix} = \beta^2 s + 2\alpha\beta t,$$

which had previously been seen in Example 2 as $df_{\mathbf{a}}(\mathbf{t})$. ■

Stress, for *scalar-valued* Fréchet differentiable functions $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ we have the important

$$\boxed{\forall \mathbf{t} \in \mathbb{R}^n,\ df_{\mathbf{a}}(\mathbf{t}) = \nabla f(\mathbf{a}) \bullet \mathbf{t}.}$$

Aside Though the gradient vector can only exist for scalar-valued functions the Jacobian matrix for vector-valued functions can be written in terms of gradient vectors. Given a vector-valued function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ then

$$\mathbf{f}(\mathbf{a}) = \begin{pmatrix} f^1(\mathbf{a}) \\ f^2(\mathbf{a}) \\ f^3(\mathbf{a}) \\ \vdots \\ f^m(\mathbf{a}) \end{pmatrix} \quad \text{so} \quad J\mathbf{f}(\mathbf{a}) = \begin{pmatrix} Jf^1(\mathbf{a}) \\ Jf^2(\mathbf{a}) \\ Jf^3(\mathbf{a}) \\ \vdots \\ Jf^m(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} \leftarrow & \nabla f^1(\mathbf{a})^T & \rightarrow \\ \leftarrow & \nabla f^2(\mathbf{a})^T & \rightarrow \\ \leftarrow & \nabla f^3(\mathbf{a})^T & \rightarrow \\ & \vdots & \\ \leftarrow & \nabla f^m(\mathbf{a})^T & \rightarrow \end{pmatrix}.$$

This could have been seen directly from (4).

2.9 *Use of the Jacobian matrix

Assume $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is Fréchet differentiable at $\mathbf{a} \in U$. Then the Jacobian matrix satisfies $d\mathbf{f}_{\mathbf{a}}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t}$ for all $\mathbf{t} \in \mathbb{R}^n$. But also, by Proposition 15, if $\mathbf{v}$ is a unit vector then $d\mathbf{f}_{\mathbf{a}}(\mathbf{v}) = d_{\mathbf{v}}\mathbf{f}(\mathbf{a})$. Combine these two facts to get

$$\boxed{\mathbf{f}\text{ F-differentiable at }\mathbf{a} \implies \forall\text{ unit }\mathbf{v},\ d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = J\mathbf{f}(\mathbf{a})\,\mathbf{v}.} \tag{7}$$

Interesting though this is, the contrapositive is very useful,

$$\boxed{\exists\text{ unit }\mathbf{v},\ d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) \neq J\mathbf{f}(\mathbf{a})\,\mathbf{v} \implies \mathbf{f}\text{ not F-differentiable at }\mathbf{a}.}$$

The argument can be repeated for scalar-valued functions $f : U \to \mathbb{R}$, starting from $df_{\mathbf{a}}(\mathbf{t}) = \nabla f(\mathbf{a}) \bullet \mathbf{t}$. This give

$$\boxed{f\text{ F-differentiable at }\mathbf{a} \implies \forall\text{ unit }\mathbf{v},\ d_{\mathbf{v}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v}.}$$

Again the contrapositive is very useful,

$$\boxed{\exists\text{ unit }\mathbf{v},\ d_{\mathbf{v}}f(\mathbf{a}) \neq \nabla f(\mathbf{a}) \bullet \mathbf{v} \implies f\text{ not F-differentiable at }\mathbf{a}.}$$

It is interesting that we can say a function is **not** Fréchet differentiable by looking at directional derivatives (remembering that $J\mathbf{f}(\mathbf{a})$ and $\nabla f(\mathbf{a})$ are defined in terms of partial derivatives, themselves directional derivatives).

Can we look at directional derivatives and deduce that a function **is** Fréchet differentiable? We know that we need more than that the directional derivatives exist, but what exactly? This is the subject of the section after next. But first we look at a situation where the directional derivative arises and it's interpretation in terms of the gradient vector gives a 'nice' result.

2.10 What can the directional derivative represent?

For a *scalar*-valued function of two variables, $f : U \subseteq \mathbb{R}^2 \to \mathbb{R}$, the *graph of* f is the set

$$G_f = \left\{ \begin{pmatrix} x \\ y \\ f\left((x,y)^T\right) \end{pmatrix} : \begin{pmatrix} x \\ y \end{pmatrix} \in U \right\} = \left\{ \begin{pmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{pmatrix} : \mathbf{x} \in U \right\}.$$

As an example of the meaning of the directional derivative imagine you are at a point $\mathbf{p}$ on the graph. What happens to your height as you move from that point? Be aware that the direction of your travel away from $\mathbf{p}$ is a vector $\mathbf{v}$ in $\mathbb{R}^2$ not $\mathbb{R}^3$ (just as when you are on a real mountain your direction is given in terms of north and east; up or down is not given).

Example 19 *In the special case of* $f : U \subseteq \mathbb{R}^2 \to \mathbb{R}$, *the directional derivative* $d_{\mathbf{v}} f(\mathbf{q})$ *for* $\mathbf{q} \in U$, *represents the rate of change in the* z*-coordinate as you move away, on the graph of* f, *from the point* $\mathbf{p} = \left(\mathbf{q}^T, f(\mathbf{q})\right)^T$ *in the direction* $\mathbf{v}$.

Verification If $\mathbf{q} \in U$ then $\mathbf{p} = \left(\mathbf{q}^T, f(\mathbf{q})\right)^T \in G_f$. If we move from $\mathbf{q}$ to $\mathbf{q} + t\mathbf{v} \in U$ then the point $\mathbf{p}$ on the graph moves to $\mathbf{p}'$ say. The change in z-coordinates of $\mathbf{p}'$ and $\mathbf{p}$ is $f(\mathbf{q}+t\mathbf{v}) - f(\mathbf{q})$ and so the rate of change in the z-coordinate is

$$\lim_{t \to 0} \frac{f(\mathbf{q}+t\mathbf{v}) - f(\mathbf{q})}{t} = d_{\mathbf{v}} f\left(\mathbf{q}\right),$$

if the limit exists. ■

In other words, as you stand at the point on the graph lying above $\mathbf{q} \in U$ and move from that point in the direction $\mathbf{v}$ then $d_{\mathbf{v}} f(\mathbf{q})$ represents how fast you ascend or descend.

Question In which directions is the quickest ascent and quickest descent?

If f is Fréchet differentiable then

$$d_{\mathbf{v}} f(\mathbf{q}) = \nabla f(\mathbf{q}) \bullet \mathbf{v} = \left|\nabla f(\mathbf{q})\right| \left|\mathbf{v}\right| \cos\theta$$

where θ is the angle between the vectors $\nabla f(\mathbf{q})$ and $\mathbf{v}$. Since $-1 \leq \cos\theta \leq 1$ we have

$$-\left|\nabla f(\mathbf{q})\right|\left|\mathbf{v}\right| \leq d_{\mathbf{v}} f(\mathbf{q}) \leq \left|\nabla f(\mathbf{q})\right|\left|\mathbf{v}\right|.$$

Thus $d_{\mathbf{v}} f(\mathbf{q})$ is maximal, i.e. we have quickest ascent, when $\cos\theta = 1$, i.e. $\theta = 0$. Hence $\max_{\mathbf{v}:|\mathbf{v}|=1} d_{\mathbf{v}} f(\mathbf{a}) = |\nabla f(\mathbf{q})|$, with the maximum occurring at $\mathbf{v} = \nabla f(\mathbf{q}) \,/\, |\nabla f(\mathbf{q})|$.

Similarly, $\min_{\mathbf{v}:|\mathbf{v}|=1} d_{\mathbf{v}} f(\mathbf{a}) = -\left|\nabla f(\mathbf{q})\right|$, the quickest descent, with the minimum occurring at $\mathbf{v} = -\nabla f(\mathbf{q}) \,/\, |\nabla f(\mathbf{q})|$.

2.11 When is a function Fréchet differentiable?

Recall

Proposition 10 *Suppose that $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a function defined on an open set U with component functions $f^i : U \to \mathbb{R}$, $1 \leq i \leq m$, and $\mathbf{a} \in U$. Then*

$$d\mathbf{f}_{\mathbf{a}} \text{ exists} \iff \forall\, 1 \leq i \leq m,\ df^i_{\mathbf{a}} \text{ exists.}$$

Further

$$d\mathbf{f}_{\mathbf{a}} \text{ exists} \implies \forall\, 1 \leq i \leq m,\ \ (d\mathbf{f}_{\mathbf{a}})^i = df^i_{\mathbf{a}}.$$

Hence we can restrict ourselves to **scalar-valued** functions for which we have the following **fundamental** result**.**

Theorem 20 *Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is an open set, is a scalar-valued function such that all the partial derivatives $d_j f : U \to \mathbb{R}$ exist* ***and are continuous on*** *U (for $1 \leq j \leq n$). Then f is Fréchet differentiable on U.*

Proof Let $\mathbf{a} \in U$ be given. Since U is an open set there exists $\delta_0 > 0$ such that the ball $B_{\delta_0}(\mathbf{a}) \subseteq U$.

Let $\varepsilon > 0$ be given. Let $1 \leq j \leq n$ be given. Then the assumption that $\partial f / \partial x^j$ is continuous at $\mathbf{a}$ means there exists $\delta_j > 0$ such that

$$|\mathbf{x} - \mathbf{a}| < \delta_i \implies \left| \frac{\partial f}{\partial x^j}(\mathbf{x}) - \frac{\partial f}{\partial x^j}(\mathbf{a}) \right| < \frac{\varepsilon}{n}. \tag{8}$$

Let $\delta = \min_{0 \le j \le n} \delta_j$. Let $\mathbf{t} \in \mathbb{R}^n$ satisfy $|\mathbf{t}| < \delta$. We wish to consider $f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a})$.

1) *We produce a piece-wise straight path from* $\mathbf{a}$ *to* $\mathbf{a}+\mathbf{t}$ *where every part of the path is parallel to some axis. For example going from* $(1,2,3)$ *to* $(5,4,2)$ *in* $\mathbb{R}^3$ *we might go from* $(1,2,3)$ *to* $(5,2,3)$ *to* $(5,4,3)$ *to* $(5,4,2)$.

In general the path will be

$$\begin{pmatrix} a^1 \\ a^2 \\ a^3 \\ a^4 \\ \vdots \\ \vdots \\ a^n \end{pmatrix} \to \begin{pmatrix} a^1+t^1 \\ a^2 \\ a^3 \\ a^4 \\ \vdots \\ \vdots \\ a^n \end{pmatrix} \to \begin{pmatrix} a^1+t^1 \\ a^2+t^2 \\ a^3 \\ a^4 \\ \vdots \\ \vdots \\ a^n \end{pmatrix} \to \begin{pmatrix} a^1+t^1 \\ a^2+t^2 \\ a^3+t^3 \\ a^4 \\ \vdots \\ \vdots \\ a^n \end{pmatrix} \to \cdots \to \begin{pmatrix} a^1+t^1 \\ a^2+t^2 \\ a^3+t^3 \\ a^4+t^4 \\ \vdots \\ a^{n-1}+t^{n-1} \\ a^n \end{pmatrix} \to \begin{pmatrix} a^1+t^1 \\ a^2+t^2 \\ a^3+t^3 \\ a^4+t^4 \\ \vdots \\ a^{n-1}+t^{n-1} \\ a^n+t^n \end{pmatrix}.$$

This can be written as

$$\mathbf{b}_0 \to \mathbf{b}_1 \to \mathbf{b}_2 \to \ldots \to \mathbf{b}_{n-1} \to \mathbf{b}_n$$

where $\mathbf{b}_0 = \mathbf{a}$, $\mathbf{b}_j - \mathbf{b}_{j-1} = t^j \mathbf{e}_j$ for some $t^j \neq 0$ and $\mathbf{b}_n = \mathbf{a}+\mathbf{t}$. thus the j-th part of this path, from $\mathbf{b}_{j-1}$ to $\mathbf{b}_j$, is parallel to the j-th basis vector $\mathbf{e}_j$.

In this way we can write

$$\begin{aligned} f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) &= f(\mathbf{b}_n) - f(\mathbf{b}_0) \\ &= \big(f(\mathbf{b}_n) - f(\mathbf{b}_{n-1})\big) + \big(f(\mathbf{b}_{n-1}) - f(\mathbf{b}_{n-2})\big) + \ldots \\ &\qquad \ldots + \big(f(\mathbf{b}_2) - f(\mathbf{b}_1)\big) + \big(f(\mathbf{b}_1) - f(\mathbf{b}_0)\big) \\ &= \sum_{j=1}^{n} \big(f(\mathbf{b}_j) - f(\mathbf{b}_{j-1})\big). \end{aligned} \tag{9}$$

2) *We now apply the Mean Value Theorem to* f *on the straight line joining the points* $\mathbf{b}_{j-1}$ *and* $\mathbf{b}_j$. *The derivative will be in the direction of* $\mathbf{b}_j - \mathbf{b}_{j-1}$, *i.e.* $\mathbf{e}_j$. *A derivative in the direction of a standard basis vector i.e. a partial derivative.*

We have seen that the difference between $\mathbf{b}_{j-1}$ and $\mathbf{b}_j$ is $t^j \mathbf{e}_j$ so the straight line joining the points $\mathbf{b}_{j-1}$ and $\mathbf{b}_j$ is $\mathbf{b}_{j-1} + s\mathbf{e}_j$, where s is a variable $0 \le s \le t^j$ (though t^j maybe negative in which case $t^j \le s \le 0$).

Define a function

$$\phi : \mathbb{R} \to \mathbb{R},\ \phi(s) = f(\mathbf{b}_{j-1} + s\,\mathbf{e}_j),$$

so

$$f(\mathbf{b}_j) - f(\mathbf{b}_{j-1}) = \phi\left(t^j\right) - \phi(0). \tag{10}$$

Yet ϕ is a real-valued function of one variable and we would like to apply results from Real Analysis, in particular the Mean Value Theorem. For this we need to know that ϕ is differentiable. Let $s : \mathbf{b}_{j-1} + s\,\mathbf{e}_j \in U$ and consider

$$\begin{aligned} \lim_{h\to 0} \frac{\phi(s+h) - \phi(s)}{h} &= \lim_{h\to 0} \frac{f(\mathbf{b}_{j-1} + (s+h)\,\mathbf{e}_j) - f(\mathbf{b}_{j-1} + s\,\mathbf{e}_j)}{h} \\ &= \lim_{h\to 0} \frac{f((\mathbf{b}_{j-1} + s\,\mathbf{e}_j) + h\mathbf{e}_j) - f(\mathbf{b}_{j-1} + s\,\mathbf{e}_j)}{h}. \end{aligned}$$

This is the definition of the directional derivative of f at $\mathbf{b}_{j-1} + s\mathbf{e}_j$ in the direction $\mathbf{e}_j$, which is the partial derivative w.r.t. x^j. This exists, by assumption, for all $s : \mathbf{b}_{j-1} + s\,\mathbf{e}_j \in U$ and thus ϕ is differentiable at s with

$$\phi'(s) = \frac{\partial f}{\partial x^j}(\mathbf{b}_{j-1} + s\,\mathbf{e}_j). \tag{11}$$

The set of s such that $\mathbf{b}_{j-1} + s\,\mathbf{e}_j \in U$ is an open interval containing the closed interval $[0, t^j]$. Since differentiable implies continuous we deduce that ϕ is continuous on $[0, t^j]$ and differentiable on $(0, t^j)$. Thus we can apply the Mean Value Theorem to ϕ on $[0, t^j]$ to deduce that

$$\phi\left(t^j\right) - \phi(0) = \phi'\left(c^j\right)\left(t^j - 0\right),$$

for some c^j between 0 and t^j (I did not say t^j was positive). Thus, by (11),

$$\phi\left(t^j\right) - \phi(0) = \phi'\left(c^j\right) t^j = \frac{\partial f}{\partial x^j}(\mathbf{w}_j)\, t^j, \tag{12}$$

where $\mathbf{w}_j = \mathbf{b}_{j-1} + c^j\mathbf{e}_j$. Think of $\mathbf{w}_j$ as a point on the straight line joining $\mathbf{b}_{j-1}$ with $\mathbf{b}_j$.

Then (9), (10) and (12) combine as

$$f(\mathbf{a} + \mathbf{t}) - f(\mathbf{a}) = \sum_{j=1}^{n} t^j \frac{\partial f}{\partial x^j}(\mathbf{w}_j). \tag{13}$$

3) *Next we use the fact that the* $d_j f$ *are all continuous at* $\mathbf{a}$. *We use this is the form* (8), *which requires showing that* $|\mathbf{w}_j - \mathbf{a}| < \delta$ *for all* $1 \leq j \leq n$.

But $\mathbf{w}_j - \mathbf{a} = \mathbf{b}_{j-1} + c^j \mathbf{e}_j - \mathbf{a}$ which equals

$$\begin{pmatrix} a^1 + t^1 \\ a^2 + t^2 \\ \vdots \\ a^{j-1} + t^{j-1} \\ a^j \\ a^{j+1} \\ \vdots \\ a^n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c^j \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} a^1 \\ a^2 \\ \vdots \\ a^{j-1} \\ a^j \\ a^{j+1} \\ \vdots \\ a^n \end{pmatrix} = \begin{pmatrix} t^1 \\ t^2 \\ \vdots \\ t^{j-1} \\ c^j \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

So

$$\begin{aligned} |\mathbf{w}_j - \mathbf{a}|^2 &= \sum_{k=1}^{j-1} \left(t^k\right)^2 + \left(c^j\right)^2 \leq \sum_{k=1}^{j} \left(t^k\right)^2 \quad \text{since } \left|c^j\right| \leq \left|t^j\right| \\ &\leq \sum_{k=1}^{n} \left(t^k\right)^2 = |\mathbf{t}|^2 . \end{aligned}$$

Yet early on we assumed $|\mathbf{t}| < \delta$ therefore $|\mathbf{w}_j - \mathbf{a}| \leq |\mathbf{t}| < \delta$. Hence, by (8),

$$\left| \frac{\partial f}{\partial x^j}(\mathbf{w}_j) - \frac{\partial f}{\partial x^j}(\mathbf{a}) \right| < \frac{\varepsilon}{n}. \tag{14}$$

Thus, by (13), we get

$$\begin{aligned} \left| f(\mathbf{a} + \mathbf{t}) - f(\mathbf{a}) - \sum_{j=1}^{n} t^j \frac{\partial f}{\partial x^j}(\mathbf{a}) \right| &= \left| \sum_{j=1}^{n} t^j \left(\frac{\partial f}{\partial x^j}(\mathbf{w}_j) - \frac{\partial f}{\partial x^j}(\mathbf{a}) \right) \right| \\ &\leq \frac{\varepsilon}{n} \sum_{j=1}^{n} \left|t^j\right| \qquad \text{by (14)} \\ &\leq \frac{\varepsilon}{n} (n\,|\mathbf{t}|) = \varepsilon\,|\mathbf{t}| . \end{aligned}$$

With the linear function

$$L(\mathbf{t}) = \sum_{j=1}^{n} t^j \frac{\partial f}{\partial x^j}(\mathbf{a}) \tag{15}$$

we therefore have, for $|\mathbf{t}| < \delta$,

$$\left| \frac{f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - L(\mathbf{t})}{|\mathbf{t}|} \right| < \varepsilon.$$

Thus we have verified the definition of

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - L(\mathbf{t})}{|\mathbf{t}|} = 0.$$

Hence we have verified the definition that f is Fréchet differentiable at $\mathbf{a}$.

Yet $\mathbf{a} \in U$ was arbitrary so f is Fréchet differentiable on U. ■

Aside The proof also gives us that

$$\begin{aligned} df_{\mathbf{a}}(\mathbf{t}) &= L(\mathbf{t}) = \sum_{j=1}^{n} t^j \frac{\partial f}{\partial x^j}(\mathbf{a}) \qquad \text{from (15)} \\ &= \nabla f(\mathbf{a}) \bullet \mathbf{t}. \end{aligned}$$

But this was already known to follow from f Fréchet differentiable.

For vector-valued functions we have

Corollary 21 *Suppose that* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, *where* U *is an open set, is a function such that all the partial derivatives* $d_j\mathbf{f} : U \to \mathbb{R}^m$ *exist* ***and are continuous on*** U *(for* $1 \leq j \leq n$*). Then* $\mathbf{f}$ *is Fréchet differentiable on* U.

Proof *Not given* The assumption of the corollary is that for all $1 \leq j \leq n$ the $d_j\mathbf{f}$ are continuous. The fact that a vector-valued function is continuous iff all the component functions are continuous justifies the first implication in

$$
\begin{aligned}
\forall 1\le j\le n,\ d_j\mathbf{f} \text{ is continuous} &\iff \forall 1\le j\le n, \forall 1\le i\le m,\ (d_j\mathbf{f})^i \text{ is continuous}\\
&\iff \forall 1\le j\le n, \forall 1\le i\le m,\ d_j f^i \text{ is continuous}\\
&\iff \forall 1\le i\le m, \forall 1\le j\le n,\ d_j f^i \text{ is continuous,}\\
&\qquad \text{on interchanging the quantifiers}\\
&\iff \forall 1\le i\le m,\ f^i \text{ is Fréchet differentiable,}\\
&\qquad \text{by Theorem 20,}\\
&\iff \mathbf{f} \text{ is Fréchet differentiable,}
\end{aligned}
$$

by Proposition 10. ■

Definition 22 *If* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, *where* U *is an open set, is continuous on* U *we say that* $\mathbf{f}$ *is a* ***function of class*** $\mathcal{C}^0$.

If the partial derivatives $d_j\mathbf{f}(\mathbf{x})$ $1 \le j \le n$ *exist for all* $\mathbf{x} \in U$ *and are continuous functions on* U *then* $\mathbf{f}$ *is a* ***function of class*** $\mathcal{C}^1$.

The result of Corollary 21 can be rephrased; if a function is of class $\mathcal{C}^1$ then it is a Fréchet differentiable function, i.e.

$$\boxed{\mathbf{f} \text{ is } \mathcal{C}^1 \implies \mathbf{f} \text{ is Fréchet differentiable.}}$$

Stress The *converse is false*, there exist Fréchet differentiable functions that are not $\mathcal{C}^1$. So

$$\boxed{\mathbf{f} \text{ is Fréchet differentiable} \;\;\not\Longrightarrow\;\; \mathbf{f} \text{ is } \mathcal{C}^1.}$$

See the Appendix for an example.

2.12 Examples

Given $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$

- Calculate the Jacobian matrix,
- Check that all entries are continuous on U,
- If they are, $\mathbf{f}$ is Fréchet differentiable on U and $d\mathbf{f}_{\mathbf{a}}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t}$.

We have seen the first example numerous times previously, but you can now see how the use of Theorem 20 is far quicker than the verification of the definition.

Example 4 **continued** *Let* $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} xy^2 \\ x^2 y \end{pmatrix},$$

for $\mathbf{x} \in \mathbb{R}^2$. *Show that* $\mathbf{f}$ *is Fréchet differentiable on* $\mathbb{R}^2$ *and find the derivative,* $d\mathbf{f}_{\mathbf{a}}(\mathbf{t})$ *for* $\mathbf{a} \in \mathbb{R}^2$.

Solution The Jacobian matrix of $\mathbf{f}$ at $\mathbf{x} \in \mathbb{R}^2$ is

$$J\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \partial f^1(\mathbf{x})/\partial x & \partial f^1(\mathbf{x})/\partial y \\ \partial f^2(\mathbf{x})/\partial x & \partial f^2(\mathbf{x})/\partial y \end{pmatrix} = \begin{pmatrix} y^2 & 2xy \\ 2xy & x^2 \end{pmatrix}.$$

All entries are polynomials and thus continuous on $\mathbb{R}^2$. Hence $\mathbf{f}$ is a C^1-function and thus Fréchet differentiable on $\mathbb{R}^2$. Further, with $\mathbf{a} = (\alpha, \beta)^T$,

$$d\mathbf{f}_{\mathbf{a}}(\mathbf{t}) = J\mathbf{f}(\mathbf{a})\,\mathbf{t} = \begin{pmatrix} \beta^2 & 2\alpha\beta \\ 2\alpha\beta & \alpha^2 \end{pmatrix} \mathbf{t} = \begin{pmatrix} \beta^2 s + 2\alpha\beta t \\ 2\alpha\beta s + \alpha^2 t \end{pmatrix}$$

for all $\mathbf{t} = (s, t)^T$, agreeing with earlier results. ■

Example 23 *Let* $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ *be the function*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \cos(xy^2) \\ \sin(x^2 y) \end{pmatrix}.$$

Show that $\mathbf{f}$ *is Fréchet differentiable on* $\mathbb{R}^3$ *and find* $J\mathbf{f}(\mathbf{x})$.

Solution The Jacobian matrix of $\mathbf{f}$ at $\mathbf{x} = (x, y)^T \in \mathbb{R}^2$ is

$$\begin{aligned} J\mathbf{f}(\mathbf{x}) &= \begin{pmatrix} d_1\mathbf{f}(\mathbf{x}) & d_2\mathbf{f}(\mathbf{x}) \end{pmatrix} \\ &= \begin{pmatrix} -y^2 \sin(xy^2) & -2xy \sin(xy^2) \\ 2xy \cos(x^2y) & x^2 \cos(x^2y) \end{pmatrix}. \end{aligned}$$

All these four functions are continuous on $\mathbb{R}^3$. Hence, by Corollary 21, $\mathbf{f}$ is Fréchet differentiable on $\mathbb{R}^3$. ■

Aside (*Not in lectures*) Continuing with the example, if $\mathbf{x}_1 = (\pi, 1)^T$ we get the Jacobian matrix

$$J\mathbf{f}(\mathbf{x}_1) = \begin{pmatrix} 0 & 0 \\ 2\pi \cos(\pi^2) & \pi^2 \cos(\pi^2) \end{pmatrix}.$$

If $\mathbf{x}_2 = (1, 1)^T$ then

$$J\mathbf{f}(\mathbf{x}_2) = \begin{pmatrix} -\sin 1 & -\sin 1 \\ 2\cos 1 & \cos 1 \end{pmatrix}.$$

A difference between $J\mathbf{f}(\mathbf{x}_1)$ and $J\mathbf{f}(\mathbf{x}_2)$ is that the former matrix has rank 1 (the number of independent rows or columns) whereas the second has rank 2 (and in fact is *full rank* since it cannot have more than 2 independent rows). The rank of the Jacobian matrix will be seen to be important later.

I have gone over the following example in the tutorials; it is an extended version of a question on Sheet 3.

Example 24 *Let*

$$f(\mathbf{x}) = \frac{x^2 y}{x^2 + y^2}$$

for $\mathbf{x} \neq \mathbf{0}$, *with* $f(\mathbf{0}) = 0$.

i. Prove that f *is continuous at* $\mathbf{0}$

ii. Calculate

$$\frac{\partial f}{\partial x}(\mathbf{0}) \quad \textit{and} \quad \frac{\partial f}{\partial y}(\mathbf{0}).$$

Further, calculate

$$\frac{\partial f}{\partial x}(\mathbf{x}) \quad \textit{and} \quad \frac{\partial f}{\partial y}(\mathbf{x})$$

for $\mathbf{x} \neq \mathbf{0}$. *Are the partial derivatives continuous at* $\mathbf{0}$*, i.e. is* f *a* $\mathcal{C}^1$*-function at* $\mathbf{0}$*?*

iii. Calculate $d_{\mathbf{v}} f(\mathbf{0})$ *for any vector* $\mathbf{v} \neq \mathbf{0}$.

iv. Is f *Fréchet differentiable at* $\mathbf{0}$*?*

Solution i. The function f is continuous at $\mathbf{0}$ if $\lim_{\mathbf{x}\to\mathbf{0}} f(\mathbf{x}) = f(\mathbf{0}) = 0$. We prove this using the Sandwich Rule.

$$|f(\mathbf{x})| = \left|\frac{x^2 y}{x^2+y^2}\right| \leq \frac{|x|^2|y|}{|\mathbf{x}|^2} \leq \frac{|\mathbf{x}|^2|\mathbf{x}|}{|\mathbf{x}|^2} = |\mathbf{x}| \to 0$$

as $\mathbf{x} \to \mathbf{0}$. Hence $f(\mathbf{x}) \to \mathbf{0}$ and so f is continuous at $\mathbf{0}$.

ii. By definition

$$\begin{aligned}\frac{\partial f}{\partial x}(\mathbf{0}) &= d_{\mathbf{e}_1} f(\mathbf{0}) = \lim_{t\to 0} \frac{f(\mathbf{0}+t\mathbf{e}_1) - f(\mathbf{0})}{t} \\ &= \lim_{t\to 0} \frac{1}{t} f\left(\begin{pmatrix} t \\ 0 \end{pmatrix}\right) = \lim_{t\to 0} \frac{t^2 \times 0}{t(t^2+0^2)} = 0.\end{aligned}$$

Similarly

$$\frac{\partial f}{\partial y}(\mathbf{0}) = 0.$$

For $\mathbf{x} \neq \mathbf{0}$ there is no need to go back to the definition but simply apply partial differentiation to get

$$\frac{\partial f}{\partial x}(\mathbf{x}) = \frac{2xy^3}{(x^2+y^2)^2} \quad \text{and} \quad \frac{\partial f}{\partial y}(\mathbf{x}) = \frac{x^4 - x^2y^2}{(x^2+y^2)^2}.$$

Are these continuous at $\mathbf{0}$? We will only look at $\partial f/\partial x$. So is

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial f}{\partial x}(\mathbf{x}) = \frac{\partial f}{\partial x}(\mathbf{0}),$$

i.e.

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{2xy^3}{(x^2+y^2)^2} = 0?$$

My guess would be no; in the rational function the total powers on the top are 4, $(3+1)$, and on the bottom also 4, (2×2). For convergence I would expect to see higher powers on the top. Believing the limit to not exist we look at different directional limits. First along the x-axis,

$$\lim_{t\to 0+}\frac{\partial f}{\partial x}(t\mathbf{e}_1)=0.$$

But along $y=x$, i.e. $\mathbf{v}=(1,1)^T/\sqrt{2}$, we find

$$\lim_{t\to 0+}\frac{\partial f}{\partial x}(t\mathbf{v})=1.$$

Hence the partial derivative is not continuous at $\mathbf{0}$ and so f is not a $\mathcal{C}^1$-function at $\mathbf{0}$.

iii. I leave it to the student to check that for non-zero $\mathbf{v}=(u,v)^T\in\mathbb{R}^2$, we have

$$d_{\mathbf{v}}f(\mathbf{0})=\frac{u^2v}{u^2+v^2}.$$

iv. Assume f Fréchet differentiable at $\mathbf{0}$. This means the Fréchet derivative gives us the directional derivative,

$$df_{\mathbf{0}}f(\mathbf{v})=d_{\mathbf{v}}f(\mathbf{0}).$$

Yet the Fréchet derivative is given by the gradient vector, so for all non-zero $\mathbf{v}\in\mathbb{R}^2$,

$$\begin{aligned} df_{\mathbf{0}}(\mathbf{v}) &= \nabla f(\mathbf{0})\bullet\mathbf{v}=\begin{pmatrix}\partial f/\partial x(\mathbf{0})\\ \partial f/\partial y(\mathbf{0})\end{pmatrix}\bullet\mathbf{v}\\ &= \begin{pmatrix}0\\0\end{pmatrix}\bullet\mathbf{v}=0,\end{aligned}$$

having used the results of Part ii. Combine we deduce $d_{\mathbf{v}}f(\mathbf{0})=0$ for all non-zero $\mathbf{v}\in\mathbb{R}^2$. Yet, if we choose $\mathbf{v}=(1,1)^T/\sqrt{2}$ in Part iii, we see that $d_{\mathbf{v}}f(\mathbf{0})=1/2^{3/2}\neq 0$. This contradiction means that f is not Fréchet differentiable at $\mathbf{0}$. ∎

2.13 Product and Quotient functions

Example 25 *Let* $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$.
i. The product function $p : \mathbb{R}^2 \to \mathbb{R}, (x,y)^T \mapsto xy$ *is Fréchet differentiable on* $\mathbb{R}^2$ *with*

$$\nabla p(\mathbf{k}) = \begin{pmatrix} \ell \\ k \end{pmatrix} \quad \textit{and} \quad dp_{\mathbf{k}}(\mathbf{t}) = \begin{pmatrix} \ell \\ k \end{pmatrix} \bullet \mathbf{t} = \ell s + kt,$$

for any $\mathbf{k} = (k,\ell)^T \in \mathbb{R}^2$.
ii. The quotient function $q : \mathbb{R} \times \mathbb{R}^\dagger, (x,y)^T \mapsto x/y$ *is Fréchet differentiable on* $\mathbb{R} \times \mathbb{R}^\dagger$ *with*

$$\nabla q(\mathbf{k}) = \frac{1}{\ell^2}\begin{pmatrix} \ell \\ -k \end{pmatrix} \quad \textit{and} \quad dq_{\mathbf{k}}(\mathbf{t}) = \frac{1}{\ell^2}\begin{pmatrix} \ell \\ -k \end{pmatrix} \bullet \mathbf{t} = \frac{\ell s - kt}{\ell^2},$$

for any $\mathbf{k} = (k,\ell)^T \in \mathbb{R} \times \mathbb{R}^\dagger$.

We have previously shown that p and q are everywhere continuous.

Proof i. The Gradient vector of p at $\mathbf{x} = (x,y)^T \in \mathbb{R}^2$ is

$$\nabla p(\mathbf{x}) = \begin{pmatrix} \partial p(\mathbf{x})/\partial x \\ \partial p(\mathbf{x})/\partial y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

The component functions are both polynomials and thus continuous on all of $\mathbb{R}^2$. Hence p is $\mathcal{C}^1$ and so, by Theorem 20, is Fréchet differentiable on $\mathbb{R}^2$. Then, for all $\mathbf{k} = (k,\ell)^T \in \mathbb{R}^2$,

$$dp_{\mathbf{k}}(\mathbf{t}) = \nabla p(\mathbf{k}) \bullet \mathbf{t} = \begin{pmatrix} \ell \\ k \end{pmatrix} \bullet \begin{pmatrix} s \\ t \end{pmatrix} = \ell s + kt,$$

for all $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$.
ii. The Gradient vector of q at $\mathbf{x} = (x,y)^T \in \mathbb{R} \times \mathbb{R}^\dagger$ is

$$\nabla q(\mathbf{x}) = \begin{pmatrix} \partial q(\mathbf{x})/\partial x \\ \partial q(\mathbf{x})/\partial y \end{pmatrix} = \begin{pmatrix} 1/y \\ -x/y^2 \end{pmatrix}.$$

The component functions are both rational functions and continuous wherever they are defined, i.e. on all of $\mathbb{R}\times\mathbb{R}^{\dagger}$. Hence q is $\mathcal{C}^1$ and so, by Theorem 20, is Fréchet differentiable on $\mathbb{R}\times\mathbb{R}^{\dagger}$. Then, for all $\mathbf{k}=(k,\ell)^T\in\mathbb{R}\times\mathbb{R}^{\dagger}$,

$$dp_{\mathbf{k}}(\mathbf{t}) = \nabla p(\mathbf{k})\bullet\mathbf{t} = \begin{pmatrix} 1/\ell \\ -k/\ell^2 \end{pmatrix}\bullet\begin{pmatrix} s \\ t \end{pmatrix} = \frac{\ell s - kt}{\ell^2},$$

for all $\mathbf{t}=(s,t)^T\in\mathbb{R}^2$. ■

Exercise for student: prove these results by verifying the definition of Fréchet differentiable. You will see how much easier it is to simply show that the given function was $\mathcal{C}^1$.

2.14 The Chain Rule

The next important situation we examine is a chain of functions

$$\mathbb{R}^p \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{\mathbf{f}} \mathbb{R}^m.$$

We have previously stated that if $\mathbf{g}$ is continuous at $\mathbf{a}\in\mathbb{R}^p$ and $\mathbf{f}$ is continuous at $\mathbf{b}=\mathbf{g}(\mathbf{a})\in\mathbb{R}^n$ then the composite $\mathbf{f}\circ\mathbf{g}$ is continuous at $\mathbf{a}$. we now see that we can replace 'is continuous' by 'is Fréchet differentiable'.

Theorem 26 ***Chain Rule*** *Suppose that* $\mathbf{g}:V\subseteq\mathbb{R}^p\to\mathbb{R}^n$ *is a function defined on an open set* V *and* $\mathbf{f}:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$ *is a function defined on an open set* U *containing the image of* $\mathbf{g}$*, i.e.* $\mathbf{g}(V)\subseteq U$*.*

If $\mathbf{g}$ *is Fréchet differentiable at* $\mathbf{a}\in V$ *and* $\mathbf{f}$ *is Fréchet differentiable at* $\mathbf{b}=\mathbf{g}(\mathbf{a})\in U$ *then the composite* $\mathbf{f}\circ\mathbf{g}:V\to\mathbb{R}^m$ *is Fréchet differentiable at* $\mathbf{a}$ *and*

$$d(\mathbf{f}\circ\mathbf{g})_{\mathbf{a}} = d\mathbf{f}_{\mathbf{b}}\circ d\mathbf{g}_{\mathbf{a}},$$

both sides linear functions from $\mathbb{R}^p$ *to* $\mathbb{R}^m$*.*

Proof not given. No new ideas over those seen in the case of scalar-valued functions of one variable, though there is an increase in detail. See Appendix.

■

A very useful form of the Chain Rule is the following,

Corollary 27 *Given the assumptions of the Chain Rule, the Jacobian matrices satisfy*

$$J(\mathbf{f} \circ \mathbf{g})\,(\mathbf{a}) = J\mathbf{f}(\mathbf{b})\,J\mathbf{g}(\mathbf{a})\,.$$

Proof The Jacobian matrix is the matrix associated with the linear map $d\,(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}}$ so, for any $\mathbf{t} \in \mathbb{R}^p$ we have

$$\begin{aligned}
J(\mathbf{f} \circ \mathbf{g})\,(\mathbf{a})\,\mathbf{t} &= d\,(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}}\,(\mathbf{t}) \\
&= (d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}})\,(\mathbf{t}) \quad \text{by the Chain Rule} \\
&= d\mathbf{f}_{\mathbf{b}}\,(d\mathbf{g}_{\mathbf{a}}\,(\mathbf{t})) \quad \text{by definition of convolution} \\
&= J\mathbf{f}(\mathbf{b})\,d\mathbf{g}_{\mathbf{a}}\,(\mathbf{t}) \quad \text{since } J\mathbf{f}(\mathbf{b}) \text{ is associated with } d\mathbf{f}_{\mathbf{b}} \\
&= J\mathbf{f}(\mathbf{b})\,J\mathbf{g}(\mathbf{a})\,\mathbf{t} \quad \text{since } J\mathbf{g}(\mathbf{b}) \text{ is associated with } d\mathbf{g}_{\mathbf{b}}.
\end{aligned}$$

True for all $\mathbf{t} \in \mathbb{R}^p$ means $J\,(\mathbf{f} \circ \mathbf{g})\,(\mathbf{a}) = J\mathbf{f}(\mathbf{b})\,J\mathbf{g}(\mathbf{a})$. ■

2.15 Chain Rule, special case

There are two special cases, $p = 1$ and $m = 1$.

If $p = 1$ then $\mathbf{g}$ and $\mathbf{f} \circ \mathbf{g}$ are **functions of one variable**, so

$$\mathbb{R} \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{\mathbf{f}} \mathbb{R}^m.$$

Recall that, for a function of one variable, we write $J\mathbf{g}(t) = \mathbf{g}'\,(t)$ and so, similarly, $J\,(\mathbf{f} \circ \mathbf{g})(t) = (\mathbf{f} \circ \mathbf{g})'\,(t)$. Thus the Chain Rule reads

$$(\mathbf{f} \circ \mathbf{g})'\,(t) = J\mathbf{f}(\mathbf{g}(t))\,\mathbf{g}'\,(t)\,. \tag{16}$$

If $m = 1$ then f and $f \circ \mathbf{g}$ are **scalar-valued**. Think of the coordinates in $\mathbb{R}^p$ as x^i for $1 \leq i \leq p$, while in $\mathbb{R}^n$ they will be y^j for $1 \leq j \leq n$. Then the Chain Rule can be written in a way you might well have seen before,

$$\frac{\partial f \circ \mathbf{g}}{\partial x^i}(\mathbf{a}) = \sum_{k=1}^{n} \frac{\partial f}{\partial y^k}(\mathbf{b})\,\frac{\partial g^k}{\partial x^i}(\mathbf{a})\,,$$

for $1 \leq i \leq p$. See Problem Sheet 5, but it follows from $J\left(f\circ\mathbf{g}\right)(\mathbf{a}) = Jf(\mathbf{b})\,J\mathbf{g}(\mathbf{a})$ in the form

$$\underbrace{\begin{pmatrix} \frac{\partial(f\circ\mathbf{g})(\mathbf{a})}{\partial x^1} & \cdots & \frac{\partial(f\circ\mathbf{g})(\mathbf{a})}{\partial x^i} & \cdots & \frac{\partial(f\circ\mathbf{g})(\mathbf{a})}{\partial x^p}\end{pmatrix}}_{i\text{-th term}} = \begin{pmatrix} \frac{\partial f(\mathbf{b})}{\partial y^1} & \cdots & \cdots & \frac{\partial f(\mathbf{b})}{\partial y^n}\end{pmatrix} \underbrace{\begin{pmatrix} \uparrow \\ | \\ | \\ \downarrow \end{pmatrix}}_{i^{\text{th}}\text{ column}}.$$

Aside 1 If $m = p = 1$, so $\mathbb{R} \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{f} \mathbb{R}$, then

$$\left(f \circ \mathbf{g}\right)'(t) = \nabla f(\mathbf{g}(t)) \bullet \mathbf{g}'(t)\,.$$

Aside 2 The case $p = 1$ was seen earlier when looking at the limits of $\mathbf{f}$ along a curve $\mathbf{g}$.

2.16 Rules for Differentiation

The following result can be proved by checking that the expressions given for the Fréchet derivatives satisfy the definition for the Fréchet derivatives of sums, products and quotients. Instead we will use the method employed to proved the Rules for Continuity, the Chain Rule.

Theorem 28 *Suppose that $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$, for an open set U, are Fréchet differentiable at $\mathbf{a} \in U$.*

a) For any $\alpha, \beta \in \mathbb{R}$ the sum $\alpha f + \beta g$ is Fréchet differentiable at $\mathbf{a}$ and

$$d\left(\alpha f + \beta g\right)_{\mathbf{a}} = \alpha df_{\mathbf{a}} + \beta dg_{\mathbf{a}}.$$

b) The product fg is Fréchet differentiable at $\mathbf{a}$ and

$$d\left(fg\right)_{\mathbf{a}} = g(\mathbf{a})\,df_{\mathbf{a}} + f(\mathbf{a})\,dg_{\mathbf{a}}.$$

c) The quotient f/g is Fréchet differentiable at $\mathbf{a}$ and

$$d\left(\frac{f}{g}\right)_{\mathbf{a}} = \frac{g(\mathbf{a})\,df_{\mathbf{a}} - f(\mathbf{a})\,dg_{\mathbf{a}}}{g(\mathbf{a})^2},$$

provided that $g(\mathbf{a}) \neq 0$.

Proof To prove equality of linear maps it suffices to prove equality of associated matrices.

$$
\begin{aligned}
J(\alpha f+\beta g)(\mathbf{a}) &= \alpha Jf(\mathbf{a})+\beta Jg(\mathbf{a})\,,\\
J(fg)(\mathbf{a}) &= g(\mathbf{a})\,Jf(\mathbf{a})+f(\mathbf{a})\,Jg(\mathbf{a})\,,\\
J(f/g)(\mathbf{a}) &= \left(g(\mathbf{a})\,Jf(\mathbf{a})-f(\mathbf{a})\,Jg(\mathbf{a})\right)/g(\mathbf{a})^2\,.
\end{aligned}
$$

The first is left to student.

For the latter two define $\mathbf{F}:U\subseteq\mathbb{R}^n\to\mathbb{R}^2$ by

$$\mathbf{F}(\mathbf{x})=\begin{pmatrix} f(\mathbf{x})\\ g(\mathbf{x})\end{pmatrix}\in\mathbb{R}^2 \tag{17}$$

for all $\mathbf{x}\in U$. Then $fg=p\circ\mathbf{F}$ where $p(\mathbf{x})=xy$ and $f/g=q\circ\mathbf{F}$ where $q(\mathbf{x})=x/y$ provided $y\neq 0$.

The component functions of $\mathbf{F}$ are, by assumption, Fréchet differentiable at $\mathbf{a}$ and thus $\mathbf{F}(\mathbf{x})$ is Fréchet differentiable with derivative

$$J\mathbf{F}(\mathbf{x})=\begin{pmatrix} Jf(\mathbf{x})\\ Jg(\mathbf{x})\end{pmatrix}. \tag{18}$$

The Chain Rule gives

$$
\begin{aligned}
J(fg)(\mathbf{a}) &= J(p\circ\mathbf{F})(\mathbf{a})=J(p(\mathbf{F}(\mathbf{a})))\,J\mathbf{F}(\mathbf{a})\,,\\
J(f/g)(\mathbf{a}) &= J(q\circ\mathbf{F})(\mathbf{a})=J(q(\mathbf{F}(\mathbf{a})))\,J\mathbf{F}(\mathbf{a})\,.
\end{aligned}
$$

From Example 25 we have

$$
\begin{aligned}
J(p(\mathbf{x})) &= (y,x) \quad\text{so}\quad J(p(\mathbf{F}(\mathbf{a})))=(g(\mathbf{a})\,,\,f(\mathbf{a}))\\
J(q(\mathbf{x})) &= (y,-x)/y^2 \quad\text{so}\quad J(q(\mathbf{F}(\mathbf{a})))=(g(\mathbf{a})\,,-f(\mathbf{a}))/g^2(\mathbf{a})\,.
\end{aligned}
$$

Thus, for the product,

$$
\begin{aligned}
J(fg)(\mathbf{a}) &= J(p(\mathbf{F}))\,J\mathbf{F}(\mathbf{a})=(g(\mathbf{a})\,,f(\mathbf{a}))\begin{pmatrix} Jf(\mathbf{a})\\ Jg(\mathbf{a})\end{pmatrix}\\
&= g(\mathbf{a})\,Jf(\mathbf{a})+f(\mathbf{a})\,Jg(\mathbf{a})\,,
\end{aligned}
$$

and, for the quotient,

$$\begin{aligned} J(f/g)\,(\mathbf{a}) &= J(q\,(\mathbf{F}))\,J\mathbf{F}(\mathbf{a}) = \frac{1}{g^2(\mathbf{a})}\,(g(\mathbf{a})\,,-f(\mathbf{a}))\begin{pmatrix} Jf(\mathbf{a}) \\ Jg(\mathbf{a}) \end{pmatrix} \\ &= (g(\mathbf{a})\,Jf(\mathbf{a}) - f(\mathbf{a})\,Jg(\mathbf{a}))\,/g(\mathbf{a})^2\,. \end{aligned}$$

■

We have previously seen the idea of introducing $\mathbf{F}$ in the proof of the limit laws for products and quotients of scalar-valued function.

2.17 *Inverse Function

Not given in lectures, though we will later come back to inverse functions.

Question what can be said if a Fréchet differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ has a Fréchet differentiable inverse?

Assume that $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ has an inverse $\mathbf{g} : V \subseteq \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{f}\,(U) \subseteq V$. Since $\mathbf{g}$ is the inverse of $\mathbf{f}$ then $\mathbf{g} \circ \mathbf{f}$ is the identity map 1_U on U. Yet 1_U is a linear map so $d\,(1_U)_{\mathbf{a}} = 1_n$, the identity map on $\mathbb{R}^n$ and thus $J1_U\,(\mathbf{a}) = I_n$, the identity matrix.

Also, since $\mathbf{g}$ is the inverse of $\mathbf{f}$ then $\mathbf{f} \circ \mathbf{g}$ is the identity map 1_V on V. For this we similarly have $d\,(1_V)_{\mathbf{a}} = 1_m$ and $J1_V\,(\mathbf{a}) = I_m$.

Corollary 29 *Suppose that $\mathbf{f} : U \to V$ is a differentiable bijection with differentiable inverse $\mathbf{g} : V \to U$ where U is non-empty and open in $\mathbb{R}^n$ and V is non-empty and open in $\mathbb{R}^m$. Let $\mathbf{a} \in U$ with $\mathbf{f}\,(\mathbf{a}) = \mathbf{b} \in V$. Then*

a) *$d\mathbf{f}_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m$ is an isomorphism (with inverse $d\mathbf{g}_{\mathbf{b}}$);*

b) *the Jacobian matrix $J\mathbf{f}\,(\mathbf{a})$ is non-singular with inverse $J\mathbf{g}\,(\mathbf{b})$;*

c) *$m = n$.*

Proof a) Since $\mathbf{g}$ is the inverse of $\mathbf{f}$, then $\mathbf{g} \circ \mathbf{f} = 1_U : U \to U$ and $\mathbf{f} \circ \mathbf{g} = 1_V : V \to V$. Hence by the Chain Rule,

$$d\mathbf{g}_{\mathbf{b}} \circ d\mathbf{f}_{\mathbf{a}} = d\,(1_U)_{\mathbf{a}} = 1_n : \mathbb{R}^n \to \mathbb{R}^n$$

and similarly

$$d\mathbf{f}_{\mathbf{a}} \circ d\mathbf{g}_{\mathbf{b}} = d\left(1_V\right)_{\mathbf{b}} = 1_m : \mathbb{R}^m \to \mathbb{R}^m.$$

That $d\mathbf{g}_{\mathbf{b}}$ is an inverse on both sides implies $d\mathbf{f}_{\mathbf{a}}$ has an inverse in which case it is an isomorphism.

b) The Chain Rule, written in matrix form, gives

$$J\mathbf{g}(\mathbf{b})\, J\mathbf{f}(\mathbf{a}) = J\left(\mathbf{g} \circ \mathbf{f}\right)(\mathbf{a}) = J1_U\left(\mathbf{a}\right) = I_n.$$

Similarly $J\mathbf{f}(\mathbf{a})\, J\mathbf{g}(\mathbf{b}) = I_m$. That $J\mathbf{g}(\mathbf{b})$ is an inverse on both sides implies $J\mathbf{f}(\mathbf{a})$ has an inverse in which case it must be non-singular.

c) The equality $m = n$ follows from either a) or b). For example, the matrix $J\mathbf{f}(\mathbf{a})$ can only be non-singular if it is square, i.e. $m = n$. ∎

This result is weak, it is possible to replace Fréchet differentiable by continuous; Suppose that $\mathbf{f} : U \to V$ is a **continuous** bijection with **continuous** inverse $\mathbf{g} : V \to U$ where U is non-empty and open in $\mathbb{R}^n$ and V is non-empty and open in $\mathbb{R}^m$. Then $m = n$.

Finally, though it is of interest to know what happens **if** a Fréchet differentiable function has a Fréchet differentiable inverse, the fundamental question must be **when** exactly does a Fréchet differentiable function have a Fréchet differentiable inverse?

This question is the subject of the next two chapters.

Appendix for Section 2

1. Partial and Directional Derivatives

It was claimed in the lectures that a function having all partial derivatives at a point is **not** sufficient to ensure the function has a directional derivative there in **all** directions.

Example 30 *Define* $f : \mathbb{R}^2 \to \mathbb{R}$ *by* $f(\mathbf{0}) = 0$ *and, for* $\mathbf{x} = (x, y)^T \neq \mathbf{0}$*,*

$$f(\mathbf{x}) = \frac{xy}{x^3 + y^3}.$$

Show that the partial derivatives $d_j f(\mathbf{a})$ *exist for all* $i = 1, 2$ *and yet* $d_{\mathbf{v}} f(\mathbf{a})$ *does not exist for* $\mathbf{v} = (1, 1)^T / \sqrt{2}$*.*

Solution For either $i = 1$ or 2 one coordinate of $t\mathbf{e}_i$ is zero and so $f(t\mathbf{e}_i) = 0$. Thus

$$d_i f(\mathbf{0}) = \lim_{t \to 0} \frac{f(\mathbf{0} + t\mathbf{e}_i) - f(\mathbf{0})}{t} = \lim_{t \to 0} 0 = 0.$$

Let $\mathbf{v} = (1, 1)^T / \sqrt{2}$ when

$$f(t\mathbf{v}) = \frac{t^2}{\sqrt{2}t^3} = \frac{1}{\sqrt{2}t}.$$

Thus

$$d_{\mathbf{v}} f(\mathbf{0}) = \lim_{t \to 0} \frac{f(\mathbf{0} + t\mathbf{v}) - f(\mathbf{0})}{t} = \lim_{t \to 0} \frac{1}{\sqrt{2}t^2}$$

which does not exist. ■

2. Directional derivative implies directional continuity

However and in whatever situation derivatives are defined we should look for a result that says that if a function is differentiable then it is continuous.

Lemma 31 Suppose that $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a function defined on an open set U and $\mathbf{a} \in U$. If $d_{\mathbf{v}}\mathbf{f}(\mathbf{a})$ exists then $\lim_{t \to 0} \mathbf{f}(\mathbf{a} + \mathbf{v}t) = \mathbf{f}(\mathbf{a})$.

Proof

$$\mathbf{f}(\mathbf{a} + \mathbf{v}t) - \mathbf{f}(\mathbf{a}) = \frac{\mathbf{f}(\mathbf{a} + \mathbf{v}t) - \mathbf{f}(\mathbf{a})}{t} \times t.$$

Then

$$\lim_{t\to 0}\left(\mathbf{f}(\mathbf{a}+\mathbf{v}t)-\mathbf{f}(\mathbf{a})\right) = \lim_{t\to 0}\left(\frac{\mathbf{f}(\mathbf{a}+\mathbf{v}t)-\mathbf{f}(\mathbf{a})}{t}\times t\right)$$

$$= \lim_{t\to 0}\frac{\mathbf{f}(\mathbf{a}+\mathbf{v}t)-\mathbf{f}(\mathbf{a})}{t}\lim_{t\to 0} t,$$

by the Product Rule for limits. This is allowable since **both** limits on the RHS exist, the first because the limit is $d_{\mathbf{v}}\mathbf{f}\,(\mathbf{a})$ which we are assuming exists. Thus

$$\lim_{t\to 0}\left(\mathbf{f}(\mathbf{a}+\mathbf{v}t)-\mathbf{f}(\mathbf{a})\right) = d_{\mathbf{v}}\mathbf{f}(\mathbf{a})\times 0 = 0.$$

3. Fréchet derivative is Unique

The following was not proved in lectures due to lack of time.

Theorem 9 *If the Fréchet derivative exists it is unique.*

Proof Assume for contradiction that there exists a function $\mathbf{f}: U\subseteq\mathbb{R}^n\to\mathbb{R}^m$, U an open subset, and a point $\mathbf{a}\in U$, at which $\mathbf{f}$ has two Fréchet derivatives, linear maps $\mathbf{L}_1$ and $\mathbf{L}_2:\mathbb{R}^n\to\mathbb{R}^m$. Then

$$\lim_{\mathbf{t}\to\mathbf{0}}\frac{\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a})-\mathbf{L}_1(\mathbf{t})}{|\mathbf{t}|}=0 \quad\text{and}\quad \lim_{\mathbf{t}\to\mathbf{0}}\frac{\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a})-\mathbf{L}_2(\mathbf{t})}{|\mathbf{t}|}=0.$$

Consider

$$\frac{\mathbf{L}_1(\mathbf{t})-\mathbf{L}_2(\mathbf{t})}{|\mathbf{t}|} = \frac{\mathbf{L}_1(\mathbf{t})-(\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a}))-\mathbf{L}_2(\mathbf{t})+(\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a}))}{|\mathbf{t}|}$$

$$= -\frac{\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a})-\mathbf{L}_1(\mathbf{t})}{|\mathbf{t}|}+\frac{\mathbf{f}(\mathbf{a}+\mathbf{t})-\mathbf{f}(\mathbf{a})-\mathbf{L}_2(\mathbf{t})}{|\mathbf{t}|}$$

$$\to \mathbf{0}$$

as $\mathbf{t}\to\mathbf{0}$.

Define $\mathbf{L}(\mathbf{t})=\mathbf{L}_1(\mathbf{t})-\mathbf{L}_2(\mathbf{t})$, a linear function. Then with $\mathbf{t}=t\mathbf{e}_j$ for any $1\le j\le n$ we see that

$$\frac{\mathbf{L}(\mathbf{t})}{|\mathbf{t}|}=\frac{\mathbf{L}(t\mathbf{e}_j)}{|t\mathbf{e}_j|}=\frac{t}{|t|}\mathbf{L}(\mathbf{e}_j)=\pm\mathbf{L}(\mathbf{e}_j)\,.$$

This can only $\to \mathbf{0}$ as $\mathbf{t} \to \mathbf{0}$ if $\mathbf{L}(\mathbf{e}_j) = \mathbf{0}$, i.e. $\mathbf{L}_1(\mathbf{e}_j) - \mathbf{L}_2(\mathbf{e}_j) = \mathbf{0}$. Hence $\mathbf{L}_1(\mathbf{e}_j) = \mathbf{L}_2(\mathbf{e}_j)$ for all $1 \leq j \leq n$. Therefore, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{L}_1(\mathbf{x}) &= \mathbf{L}_1\left(\sum_{j=1}^n x^j \mathbf{e}_j\right) = \sum_{j=1}^n x^j \mathbf{L}_1(\mathbf{e}_j) \\
&= \sum_{j=1}^n x^j \mathbf{L}_2(\mathbf{e}_j) \\
&= \mathbf{L}_2\left(\sum_{j=1}^n x^j \mathbf{e}_j\right) = \mathbf{L}_2(\mathbf{x}).
\end{aligned}$$

Hence $\mathbf{L}_1 = \mathbf{L}_2$. ■

4. Theorem 20 and the Mean Value Theorem

What makes this proof of Theorem 20 long and difficult? Perhaps its the proof of (13)? Instead of a number of applications of the Mean Value Theorem could we not have just one?

In all the following results U is an open set in $\mathbb{R}^n$, such that if $\mathbf{x}, \mathbf{y} \in U$ then the straight line between $\mathbf{x}$ and $\mathbf{y}$ also lies in U. This is the definition that U is a *convex set.*

Theorem 32 *Assume f is a scalar-valued Fréchet differentiable function, $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, where U is an open convex subset. Given distinct points $\mathbf{x}, \mathbf{y} \in U$ there exists $\mathbf{w}$, a point on the line between $\mathbf{x}$ and $\mathbf{y}$, such that*

$$f(\mathbf{y}) - f(\mathbf{x}) = df_{\mathbf{w}}(\mathbf{y} - \mathbf{x}).$$

Also

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla f(\mathbf{w}) \bullet (\mathbf{y} - \mathbf{x}) = \sum_{i=1}^n \frac{\partial f(\mathbf{w})}{\partial x^i}\left(y^i - x^i\right). \tag{19}$$

Proof Define $\psi(s) = f(\mathbf{x} + s(\mathbf{y} - \mathbf{x}))$, so $\psi(0) = f(\mathbf{x})$ and $\psi(1) = f(\mathbf{y})$. Since f is Fréchet differentiable it is continuous and thus $\psi(s)$ is too. We need show it is differentiable w.r.t s (by the definition of last semester fro functions of one variable). To this end assume $0 < s < 1$. Then

$$\begin{aligned}
\lim_{h\to 0}\frac{\psi(s+h)-\psi(s)}{h} &= \lim_{h\to 0}\frac{f\big(\mathbf{x}+(s+h)\,(\mathbf{y}-\mathbf{x})\big)-f\big(\mathbf{x}+s\,(\mathbf{y}-\mathbf{x})\big)}{h}\\
&= \lim_{h\to 0}\frac{f\big(\mathbf{w}+h\,(\mathbf{y}-\mathbf{x})\big)-f(\mathbf{w})}{h},
\end{aligned}$$

say, where $\mathbf{w}=\mathbf{x}+s\,(\mathbf{y}-\mathbf{x})=(1-s)\,\mathbf{x}+s\mathbf{y}$. This is almost a directional derivative but $\mathbf{y}-\mathbf{x}$ is not a unit vector. Let $\mathbf{v}=(\mathbf{y}-\mathbf{x})\,/\,|\mathbf{y}-\mathbf{x}|$ be the unit vector and replace h by $\eta=|\mathbf{y}-\mathbf{x}|\,h$, which also tends to 0 as $h\to 0$. Then

$$\lim_{h\to 0}\frac{\psi(s+h)-\psi(s)}{h}=|\mathbf{y}-\mathbf{x}|\lim_{\eta\to 0}\frac{f(\mathbf{w}+\eta\mathbf{v})-f(\mathbf{w})}{\eta}=|\mathbf{y}-\mathbf{x}|\,d_{\mathbf{v}}f(\mathbf{w})\,,$$

where the limit exists since f is Fréchet differentiable. Thus $\psi'(s)$ exists for all $0 < s < 1$ and satisfies

$$\psi'(s)=|\mathbf{y}-\mathbf{x}|\,d_{\mathbf{v}}f(\mathbf{w})=|\mathbf{y}-\mathbf{x}|\,\nabla f(\mathbf{w})\bullet\mathbf{v}=\nabla f(\mathbf{w})\bullet(\mathbf{y}-\mathbf{x})\,. \qquad (20)$$

We can now apply the Mean Value Theorem for scalar-valued functions of one-variable to $\psi(s)$ on $[0,1]$ and find $c: 0 < c < 1$ and

$$\begin{aligned}
f(\mathbf{y})-f(\mathbf{x}) &= \psi(1)-\psi(0)=\psi'(c)\\
&= \nabla f(\mathbf{w})\bullet(\mathbf{y}-\mathbf{x})
\end{aligned}$$

by (20), where now $\mathbf{w}=\mathbf{x}+c\,(\mathbf{y}-\mathbf{x})$. ■

The conclusion in (19) can be compared with (13) and would appear to be stronger, with only one unknown, $\mathbf{w}$, instead of n unknowns, $\mathbf{w}_i$. UNFORTUNATELY, to prove (19) we need to **assume** that f is differentiable, whereas (13) is **used** to prove that f is differentiable.

5. Fréchet differentiable does not imply $\mathcal{C}^1$.

In the lectures we showed that "f is $\mathcal{C}^1 \Rightarrow f$ is Fréchet differentiable" and stated that the converse is not true. Here we give an example of a function that is Fréchet differentiable at a point yet whose partial derivatives exist but are not continuous at that point.

Recall that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable (in the sense from last Semester) then it is Fréchet differentiable with $df_a(t) = f'(a)\, t$. So we need only find an example of a function differentiable at some point a for which f' is not continuous at a. An example can be constructed from the fact that neither

$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right) \quad \text{nor} \quad \lim_{x\to 0} \cos\left(\frac{1}{x}\right)$$

exist. The proof is by contradiction, we can find a sequence $x_n \to 0$ as $n \to \infty$ for which $\sin(1/x_n) = 1$ and another sequence $y_n \to 0$ for which $\sin(y_n) = -1$. Thus any limit value would have to lie close to both 1 and -1, an impossibility.

Example 33 *Let $f : \mathbb{R} \to \mathbb{R}$ be given by*

$$f(x) = \begin{cases} x^2 \sin\left(\dfrac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is differentiable on $\mathbb{R}$ yet f' is not continuous at $x = 0$.

Solution If $x \neq 0$ we can simply use the rules of differentiation to see that for such x

$$f'(x) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right). \tag{21}$$

When $x = 0$ return to the definition of the derivative as a limit. So

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

We **cannot** use the Product Rule on this since $\lim_{x\to 0} \sin(1/x)$ does not exist. Instead use the Sandwich Rule:

$$\left| x \sin\left(\frac{1}{x}\right) \right| \leq |x| \to 0$$

as $x \to 0$. Hence $\lim_{x\to 0} x \sin(1/x) = 0$.

Thus

$$f'(x) = \begin{cases} 2x \sin(1/x) - \cos(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

The derivative is continuous at 0 iff $\lim_{x\to 0} f'(x) = f'(0)$. Yet $\lim_{x\to 0} f'(x)$ does not exist since $\lim_{x\to 0} \cos(1/x)$. Hence f' is not continuous at $x = 0$.■

You may be disappointed that this is not an example of a function of several variables. Perhaps try $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} |\mathbf{x}|^2 \sin\left(\frac{1}{|\mathbf{x}|}\right) & \text{if } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{x} = \mathbf{0}. \end{cases}$$

If $\mathbf{x} = (x, y)^T$ then

$$\frac{\partial}{\partial x} |\mathbf{x}| = \frac{x}{|\mathbf{x}|}.$$

for $\mathbf{x} \neq \mathbf{0}$. For such $\mathbf{x}$,

$$\frac{\partial}{\partial x} f(\mathbf{x}) = 2x \sin\left(\frac{1}{|\mathbf{x}|}\right) - \frac{x}{|\mathbf{x}|} \cos\left(\frac{1}{|\mathbf{x}|}\right).$$

To see if this is continuous at $\mathbf{x} = \mathbf{0}$ look at the directional limit at $\mathbf{0}$ in the directions $\mathbf{e}_1 = (1, 0)^T$. For $t > 0$,

$$\begin{aligned} \frac{\partial}{\partial x} f(\mathbf{x}) &= 2t \sin\left(\frac{1}{|t\mathbf{e}_1|}\right) - \frac{t}{|t\mathbf{e}_1|} \cos\left(\frac{1}{|t\mathbf{e}_1|}\right) \\ &= 2t \sin\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right) \end{aligned}$$

This is the same as (21) and, as there, has no limit as $t \to 0+$. Thus f is not continuous at the origin.

You may be disappointed that this is not an example of a vector-valued function. But, for example, take $f : \mathbb{R}^2 \to \mathbb{R}$ of the last example and define $\widehat{\mathbf{f}} : \mathbb{R}^2 \to \mathbb{R}^m$ by $\widehat{f}^1 = f$, $\widehat{f}^i = 0$ for all $2 \le i \le m$.

6. Example 25 revisited

The following proof by verification of a definition was left to the student.

Example 25 *Let* $q : \mathbb{R} \times \mathbb{R}^\dagger \to \mathbb{R}$,

$$q\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \frac{x}{y}.$$

Show, by verifying the definition, that q *is Fréchet differentiable on* $\mathbb{R}\times\mathbb{R}^{\dagger}$ *and find* $dq_{\mathbf{k}}(\mathbf{t})$ *for all* $\mathbf{k}\in\mathbb{R}\times\mathbb{R}^{\dagger}$, $\mathbf{t}\in\mathbb{R}^{2}$.

Solution For $\mathbf{k}=(k,\ell)^{T}$ with $\ell\neq 0$, and $\mathbf{t}=(s,t)^{T}\in\mathbb{R}^{2}$,

$$q(\mathbf{k}+\mathbf{t})-q(\mathbf{k})=\frac{k+s}{\ell+t}-\frac{k}{\ell}=\frac{\ell s-kt}{\ell(\ell+t)}.$$

It might be difficult to guess the linear approximation to this rational function but think of $\mathbf{t}$ and thus both s and t as small. Then $1/\ell$ is the first order approximation to $1/(\ell+t)$ and so we might guess the correct linear function of s and t is

$$L_{\mathbf{k}}(\mathbf{t})=\frac{\ell s-kt}{\ell^{2}}.$$

Then, with the given $L_{\mathbf{k}}(\mathbf{t})$,

$$\begin{aligned}\frac{q(\mathbf{k}+\mathbf{t})-q(\mathbf{k})-L_{\mathbf{k}}(\mathbf{t})}{|\mathbf{t}|} &= \frac{1}{|\mathbf{t}|}\left(\frac{\ell s-kt}{\ell(\ell+t)}-\frac{\ell s-kt}{\ell^{2}}\right)\\ &= -\frac{\ell s-kt}{|\mathbf{t}|}\frac{t}{\ell^{2}(\ell+t)}.\end{aligned}$$

Since we will be letting $\mathbf{t}\to\mathbf{0}$ first demand $|\mathbf{t}|\leq|\ell|/2$. For then, by the triangle inequality and the fact that $|t|\leq|\mathbf{t}|$,

$$|\ell+t|\geq|\ell|-|t|\geq|\ell|-|\mathbf{t}|\geq|\ell|-|\ell|/2=|\ell|/2.$$

Thus

$$\left|\frac{\ell s-kt}{|\mathbf{t}|}\frac{t}{\ell^{2}(\ell+t)}\right|\leq\frac{2|\ell s-kt||t|}{|\ell|^{3}|\mathbf{t}|}\leq\frac{2(|\ell||s|+|k||t|)|t|}{|\ell|^{3}|\mathbf{t}|}.$$

To continue, use $|s|,|t|\leq|\mathbf{t}|$, so

$$\left|-\frac{\ell s-kt}{|\mathbf{t}|}\frac{t}{\ell^{2}(\ell+t)}\right|\leq\frac{2(|\ell|+|k|)}{|\ell|^{3}}|\mathbf{t}|\to 0$$

as $\mathbf{t}\to\mathbf{0}$. So again q is Fréchet differentiable at $\mathbf{a}$ with $dq_{\mathbf{k}}(\mathbf{t})=L_{\mathbf{k}}(\mathbf{t})$. ■

7. Rules of Differentiation

The following are parts of Theorem 28.

i) *If* $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ *are Fréchet differentiable at* $\mathbf{a} \in U$ *then so is the product* fg *and*

$$d\,(fg)_{\mathbf{a}} = g(\mathbf{a})\,df_{\mathbf{a}} + f(\mathbf{a})\,dg_{\mathbf{a}}.$$

ii) *If* $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ *are Fréchet differentiable at* $\mathbf{a} \in U$ *and* $g\,(\mathbf{a}) \neq 0$, *then the quotient* f/g *is differentiable and*

$$d\,(f/g)_{\mathbf{a}} = \frac{g(\mathbf{a})\,df_{\mathbf{a}} - f(\mathbf{a})\,dg_{\mathbf{a}}}{g(\mathbf{a})^2}.$$

In lectures we gave proofs based on the Chain Rule. Here we give direct proofs but we first state a result from Section 1 for scalar-valued linear functions.

Lemma 34 *If* $L : \mathbb{R}^n \to \mathbb{R}$ *is a linear map then there exists a positive constant* C *(depending on* L*) such that* $|L\,(\mathbf{t})| \leq C\,|\mathbf{t}|$ *for all* $\mathbf{t} \in \mathbb{R}^n$.

Thus, given $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ Fréchet differentiable at $\mathbf{a} \in U$ there exist constants c_f and c_g (depending also on $\mathbf{a}$) such that $|df_{\mathbf{a}}\,(\mathbf{u})| \leq c_f\,|\mathbf{u}|$ and $|df_{\mathbf{a}}\,(\mathbf{u})| \leq c_f\,|\mathbf{u}|$ for all $\mathbf{u} \in U$.

Proof i With $\mathbf{a}$, $\mathbf{a} + \mathbf{t} \in U$ consider

$$R(\mathbf{t}) = f(\mathbf{a} + \mathbf{t})\,g(\mathbf{a} + \mathbf{t}) - f(\mathbf{a})\,g(\mathbf{a}) - \big(g(\mathbf{a})\,df_{\mathbf{a}}(\mathbf{t}) + f(\mathbf{a})\,dg_{\mathbf{a}}(\mathbf{t})\big). \qquad (22)$$

To show $d\,(fg)_{\mathbf{a}} = g\,(\mathbf{a})\,df_{\mathbf{a}} + f(\mathbf{a})\,dg_{\mathbf{a}}$ it suffices to show that $R\,(\mathbf{t})\,/\,|\mathbf{t}| \to \mathbf{0}$ as $|\mathbf{t}| \to \mathbf{0}$.

Since we are assuming f is Fréchet differentiable at $\mathbf{a}$ we know something of $f(\mathbf{a}+\mathbf{t})-f(\mathbf{a})-df_{\mathbf{a}}(\mathbf{t})$, so we rearrange (22) to include this term. Thus

$$\begin{aligned}
R(\mathbf{t}) &= \big(f(\mathbf{a}+\mathbf{t})-f(\mathbf{a})-df_{\mathbf{a}}(\mathbf{t})\big)g(\mathbf{a}+\mathbf{t}) \\
&\quad +f(\mathbf{a})\,g(\mathbf{a}+\mathbf{t})+df_{\mathbf{a}}(\mathbf{t})\,g(\mathbf{a}+\mathbf{t})-f(\mathbf{a})\,g(\mathbf{a}) \\
&\quad -\big(g(\mathbf{a})\,df_{\mathbf{a}}(\mathbf{t})+f(\mathbf{a})\,dg_{\mathbf{a}}(\mathbf{t})\big) \\
&= \big(f(\mathbf{a}+\mathbf{t})-f(\mathbf{a})-df_{\mathbf{a}}(\mathbf{t})\big)g(\mathbf{a}+\mathbf{t}) \\
&\quad +\big(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})-dg_{\mathbf{a}}(\mathbf{t})\big)f(\mathbf{a}) \\
&\quad +df_{\mathbf{a}}(\mathbf{t})\,\big(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})\big).
\end{aligned} \tag{23}$$

Divide both sides by $|\mathbf{t}|$, $\mathbf{t}\neq\mathbf{0}$, and let $\mathbf{t}\to\mathbf{0}$.

By assumption both

$$\frac{f(\mathbf{a}+\mathbf{t})-f(\mathbf{a})-df_{\mathbf{a}}(\mathbf{t})}{|\mathbf{t}|},\quad \frac{g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})-dg_{\mathbf{a}}(\mathbf{t})}{|\mathbf{t}|}\to 0,$$

as $\mathbf{t}\to\mathbf{0}$. This leaves the third term, (23) , which is bounded as

$$\left|\frac{df_{\mathbf{a}}(\mathbf{t})\,\big(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})\big)}{|\mathbf{t}|}\right| < \frac{c_f\,|\mathbf{t}|\,|g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})|}{|\mathbf{t}|},$$

by Lemma 34. But since g is Fréchet differentiable at $\mathbf{a}$ it is continuous at $\mathbf{a}$, i.e. $g(\mathbf{a}+\mathbf{t})\to g(\mathbf{a})$ as $\mathbf{t}\to\mathbf{0}$. Thus

$$\left|\frac{df_{\mathbf{a}}(\mathbf{t})\,(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a}))}{|\mathbf{t}|}\right| \le c_f\,|g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})|\to 0$$

as $\mathbf{t}\to\mathbf{0}$. Hence

$$\lim_{\mathbf{t}\to\mathbf{0}}\frac{f(\mathbf{a}+\mathbf{t})\,g(\mathbf{a}+\mathbf{t})-f(\mathbf{a})\,g(\mathbf{a})-\big(g(\mathbf{a})\,df_{\mathbf{a}}(\mathbf{t})+f(\mathbf{a})\,dg_{\mathbf{a}}(\mathbf{t})\big)}{|\mathbf{t}|}=0.$$

Therefore fg is Fréchet differentiable at $\mathbf{a}$ with Fréchet derivative $d\,(fg)_{\mathbf{a}} = g(\mathbf{a})\,df_{\mathbf{a}}+f(\mathbf{a})\,dg_{\mathbf{a}}$.

ii It suffices to prove that if $g:U\to\mathbb{R}$ is Fréchet differentiable at $\mathbf{a}\in U$ and $g\,(\mathbf{a})\neq 0$, then the quotient $1/g$ is Fréchet differentiable and

$$d\left(\frac{1}{g}\right)_{\mathbf{a}} = -\frac{dg_{\mathbf{a}}}{g(\mathbf{a})^2}.$$

The general result follows using Part **i**.

With $\mathbf{a}$, $\mathbf{a}+\mathbf{t}\in U$ consider

$$\frac{1}{g(\mathbf{a}+\mathbf{t})}-\frac{1}{g(\mathbf{a})}+\frac{dg_{\mathbf{a}}(\mathbf{t})}{g(\mathbf{a})^2}. \tag{24}$$

You should not forget to note that because $g(\mathbf{a})\neq 0$ and g is continuous at $\mathbf{a}$ then $g(\mathbf{a}+\mathbf{t})\neq 0$ for $\mathbf{t}$ sufficiently small. Thus, for such $\mathbf{t}$, (24) is well-defined. We can thus look at the inverses. Rearranging, (24) equals

$$\frac{g(\mathbf{a})^2-g(\mathbf{a})\,g(\mathbf{a}+\mathbf{t})+dg_{\mathbf{a}}(\mathbf{t})\,g(\mathbf{a}+\mathbf{t})}{g(\mathbf{a}+\mathbf{t})\,g(\mathbf{a})^2}.$$

The numerator here equals

$$-g(\mathbf{a})\left(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})-dg_{\mathbf{a}}(\mathbf{t})\right)+dg_{\mathbf{a}}(\mathbf{t})\left(g(\mathbf{a}+\mathbf{t})-g(\mathbf{a})\right).$$

Dividing by $|\mathbf{t}|$, the first term $\to 0$ as $\mathbf{t}\to\mathbf{0}$ since g is differentiable at $\mathbf{a}$, the second term $\to 0$ as $\mathbf{t}\to\mathbf{0}$ by the argument seen above. Hence

$$\lim_{\mathbf{t}\to\mathbf{0}}\frac{1}{|\mathbf{t}|}\left(\frac{1}{g(\mathbf{a}+\mathbf{t})}-\frac{1}{g(\mathbf{a})}+\frac{dg_{\mathbf{a}}(\mathbf{t})}{g(\mathbf{a})^2}\right)=0.$$

Therefore $1/g$ is Fréchet differentiable at $\mathbf{a}$ with derivative $-dg_{\mathbf{a}}(\mathbf{t})/g(\mathbf{a})^2$.■

8. The Chain Rule

The most important result unproved in the notes concerned the situation

$$\mathbb{R}^p\xrightarrow{\mathbf{g}}\mathbb{R}^n\xrightarrow{\mathbf{f}}\mathbb{R}^m.$$

For the following proof it helps to use the following equivalent definition of Fréchet differentiable.

Definition 35 *The vector-valued function* $\mathbf{f}:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$ *is* ***Fréchet differentiable at*** $\mathbf{a}\in U$ *if, and only if, there exists a linear map* $d\mathbf{f}_{\mathbf{a}}:\mathbb{R}^n\to\mathbb{R}^m$ *such that*

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{a})-d\mathbf{f}_{\mathbf{a}}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|}=\mathbf{0}. \tag{25}$$

Theorem 26 Chain Rule *Suppose that* $\mathbf{g} : V \subseteq \mathbb{R}^p \to \mathbb{R}^n$ *is a function defined on an open set* V *and* $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is a function defined on an open set* U *containing the image of* $\mathbf{g}$, *i.e.* $\mathbf{g}(V) \subseteq U$.

If $\mathbf{g}$ *is Fréchet differentiable at* $\mathbf{a} \in V$ *and* $\mathbf{f}$ *is Fréchet differentiable at* $\mathbf{b} = \mathbf{g}(\mathbf{a}) \in V$ *then the composite* $\mathbf{f} \circ \mathbf{g} : V \to \mathbb{R}^m$ *is Fréchet differentiable at* $\mathbf{a}$ *and*

$$d(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}} = d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}},$$

both sides linear functions from $\mathbb{R}^p$ *to* $\mathbb{R}^m$.

Proof It suffices to prove that, given $\mathbf{a} \in V$,

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{g}(\mathbf{x})) - \mathbf{f}(\mathbf{g}(\mathbf{a})) - d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|} = \mathbf{0}. \tag{26}$$

For this we will show that $d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}}$ is *a* linear function satisfying the definition yet, by definition, $d(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}}$ is *the* unique linear function, hence $d(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}} = d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}}$.

For $\mathbf{y} \in U$ define the function

$$\mathbf{F}_{\mathbf{b}}(\mathbf{y}) = \begin{cases} \dfrac{\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{b}) - d\mathbf{f}_{\mathbf{b}}(\mathbf{y}-\mathbf{b})}{|\mathbf{y}-\mathbf{b}|} & \text{if } \mathbf{y} \neq \mathbf{b}, \\ \mathbf{0} & \text{if } \mathbf{y} = \mathbf{b}. \end{cases}$$

Since $\mathbf{f}$ is Fréchet differentiable at $\mathbf{b}$ we have $\lim_{\mathbf{y}\to\mathbf{b}} \mathbf{F}_{\mathbf{b}}(\mathbf{y}) = \mathbf{0} = \mathbf{F}_{\mathbf{b}}(\mathbf{0})$ and so $\mathbf{F}_{\mathbf{b}}$ is continuous at $\mathbf{y} = \mathbf{b}$.

Rearrange the definition to get

$$\mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{b}) + d\mathbf{f}_{\mathbf{b}}(\mathbf{y}-\mathbf{b}) + |\mathbf{y}-\mathbf{b}|\,\mathbf{F}_{\mathbf{b}}(\mathbf{y}),$$

for all $\mathbf{y} \in U$.

Apply this with $\mathbf{y} = \mathbf{g}(\mathbf{x})$ for $\mathbf{x} \in V$. Then, since $\mathbf{b} = \mathbf{g}(\mathbf{a})$,

$$\mathbf{f}\big(\mathbf{g}(\mathbf{x})\big) = \mathbf{f}\big(\mathbf{g}(\mathbf{a})\big) + d\mathbf{f}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a})\big) + |\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a})|\,\mathbf{F}_{\mathbf{b}}(\mathbf{g}(\mathbf{x})).$$

Apply this within the left hand side of (26),

$$\begin{aligned} \frac{\mathbf{f}\big(\mathbf{g}(\mathbf{x})\big) - \mathbf{f}\big(\mathbf{g}(\mathbf{a})\big) - d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|} &= \frac{d\mathbf{f}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a})\big) - d\mathbf{f}_{\mathbf{b}} \circ d\mathbf{g}_{\mathbf{a}}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|} \\ &\quad + \frac{|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a})|}{|\mathbf{x}-\mathbf{a}|}\mathbf{F}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x})\big). \end{aligned} \tag{27}$$

In the first term on the right hand side

$$\begin{aligned}
\frac{d\mathbf{f_b}\big(\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})\big)-d\mathbf{f_b}\circ d\mathbf{g_a}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|} &= \frac{d\mathbf{f_b}\big(\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})\big)-d\mathbf{f_b}\big(d\mathbf{g_a}(\mathbf{x}-\mathbf{a})\big)}{|\mathbf{x}-\mathbf{a}|} \\
&= d\mathbf{f_b}\left(\frac{\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})-d\mathbf{g_a}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|}\right),
\end{aligned}$$

since $d\mathbf{f_b}$ is a linear function. But further, since $d\mathbf{f_b}$ is linear it is continuous and so

$$\begin{aligned}
\lim_{\mathbf{x}\to\mathbf{a}} d\mathbf{f_b}\left(\frac{\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})-d\mathbf{g_a}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|}\right) &= d\mathbf{f_b}\left(\lim_{\mathbf{x}\to\mathbf{a}}\frac{\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})-d\mathbf{g_a}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|}\right) \\
&= d\mathbf{f_b}(\mathbf{0}) \qquad \text{since } \mathbf{g} \text{ is differentiable at } \mathbf{a}, \\
&= \mathbf{0}, \qquad (28)
\end{aligned}$$

since $d\mathbf{f_b}$ is linear.

For the second term on the right hand side of (27) , we make use of

Lemma 36 *If* $\mathbf{g}: V \subseteq \mathbb{R}^p \to \mathbb{R}^n$ *if differentiable at* $\mathbf{a} \in V$ *then there exists* $\delta > 0$ *and* $C > 0$ *such that*

$$|\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})| \le C\,|\mathbf{x}-\mathbf{a}|$$

for all $\mathbf{x} : |\mathbf{x}-\mathbf{a}| < \delta$.

Proof Choose $\varepsilon = 1$ in the definition of Fréchet differentiable (25) to find $\delta > 0$ such that if $|\mathbf{x}-\mathbf{a}| < \delta$ then

$$\left|\frac{\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})-d\mathbf{g_a}(\mathbf{x}-\mathbf{a})}{|\mathbf{x}-\mathbf{a}|}\right| < 1.$$

Thus

$$|\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})| < |d\mathbf{g_a}(\mathbf{x}-\mathbf{a})| + |\mathbf{x}-\mathbf{a}|.$$

Yet $d\mathbf{g_a}(\mathbf{t})$ is a linear function and we know that for any linear function there exists a constant such that $|d\mathbf{g_a}(\mathbf{t})| < \kappa\,|\mathbf{t}|$. This all combines to give the stated result with $C = \kappa + 1$. ■

Then, for $|\mathbf{x}-\mathbf{a}|<\delta$,

$$\frac{|\mathbf{g}\left(\mathbf{x}\right)-\mathbf{g}\left(\mathbf{a}\right)|}{|\mathbf{x}-\mathbf{a}|}\mathbf{F}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x})\big) \leq C\mathbf{F}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x})\big).$$

Yet we have seen that $\mathbf{F}_{\mathbf{b}}$ is continuous at $\mathbf{b}$, so by the Composite Rule for continuous functions,

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x})\big) = \mathbf{F}_{\mathbf{b}}\Big(\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x})\Big) = \mathbf{F}_{\mathbf{b}}(\mathbf{b}) = \mathbf{0}.$$

Thus

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{|\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{a})|}{|\mathbf{x}-\mathbf{a}|}\mathbf{F}_{\mathbf{b}}\big(\mathbf{g}(\mathbf{x})\big) = \mathbf{0}. \tag{29}$$

Combining (28) and (29) gives (26) as required. ■

From a dislike of not using material I have written up, I include a second proof of the Chain Rule.

Alternative Proof of Chain Rule This is a different proof to that presented in the notes; having written it up I didn't want to waste the effort. But first, recall

Lemma *If* $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ *is a linear map then there exists a positive constant* C *(depending on* L*) such that* $|\mathbf{L}\left(\mathbf{t}\right)|\leq C\left|\mathbf{t}\right|$ *for all* $\mathbf{t}\in\mathbb{R}^n$.

Proof of Chain Rule For $\mathbf{w}:\mathbf{a}+\mathbf{w}\in\mathbf{V}$ define

$$R(\mathbf{w}) = \left(\mathbf{f}\circ\mathbf{g}\right)\left(\mathbf{a}+\mathbf{w}\right) - \left(\mathbf{f}\circ\mathbf{g}\right)\left(\mathbf{a}\right) - \left(d\mathbf{f}_{\mathbf{b}}\circ d\mathbf{g}_{\mathbf{a}}\right)\left(\mathbf{w}\right). \tag{30}$$

To show that $d\mathbf{f}_{\mathbf{b}}\circ d\mathbf{g}_{\mathbf{a}}$ is the Fréchet derivative of $\mathbf{f}\circ\mathbf{g}$ it suffices to show that $R\left(\mathbf{w}\right)/\left|\mathbf{w}\right|\to\mathbf{0}$ as $\mathbf{w}\to 0$.

1) *Rearrangement of* $R(\mathbf{w})/|\mathbf{w}|$.

Write

$$r_1(\mathbf{w}) = \mathbf{g}(\mathbf{a}+\mathbf{w}) - \mathbf{g}(\mathbf{a}) - d\mathbf{g}_{\mathbf{a}}(\mathbf{w}),$$

for $\mathbf{w}:\mathbf{a}+\mathbf{w}\in V$. The assumption that $\mathbf{g}$ is Fréchet differentiable at $\mathbf{a}$ implies $r_1\left(\mathbf{w}\right)/\left|\mathbf{w}\right|\to 0$ as $\mathbf{w}\to\mathbf{0}$.

Similarly write

$$r_2(\mathbf{u}) = \mathbf{f}(\mathbf{b}+\mathbf{u}) - \mathbf{f}(\mathbf{b}) - d\mathbf{f}_{\mathbf{b}}(\mathbf{u})\,,$$

for $\mathbf{u} : \mathbf{b}+\mathbf{u} \in U$. The assumption that $\mathbf{f}$ is Fréchet differentiable at $\mathbf{b}$ implies $r_2\left(\mathbf{u}\right) / \left|\mathbf{u}\right| \to 0$ as $\mathbf{u} \to \mathbf{0}$.

Motivated by the $(\mathbf{f} \circ \mathbf{g})\left(\mathbf{a}+\mathbf{w}\right)$ term in (30) rearrange the definition of r_1 as

$$\mathbf{g}(\mathbf{a}+\mathbf{w}) = \mathbf{g}(\mathbf{a}) + d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1\left(\mathbf{w}\right).$$

Apply $\mathbf{f}$ to both sides:

$$\begin{aligned} (\mathbf{f}\circ\mathbf{g})\left(\mathbf{a}+\mathbf{w}\right) &= \mathbf{f}\big(\mathbf{g}(\mathbf{a}+\mathbf{w})\big) = \mathbf{f}\big(\mathbf{g}(\mathbf{a}) + d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big) \\ &= \mathbf{f}\big(\mathbf{b} + d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big), \end{aligned} \tag{31}$$

since $\mathbf{b} = \mathbf{g}\left(\mathbf{a}\right)$. Rearrange the definition of r_2 as

$$\mathbf{f}(\mathbf{b}+\mathbf{u}) = \mathbf{f}(\mathbf{b}) + d\mathbf{f}_{\mathbf{b}}(\mathbf{u}) + r_2(\mathbf{u})\,.$$

Motivated by (31) apply this with $\mathbf{u} = d\mathbf{g}_{\mathbf{a}}\left(\mathbf{w}\right) + r_1\left(\mathbf{w}\right)$ getting

$$\begin{aligned} \mathbf{f}\big(\mathbf{b} + d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big) &= \mathbf{f}\big(\mathbf{b}\big) + d\mathbf{f}_{\mathbf{b}}\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1\left(\mathbf{w}\right)\big) + r_2\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big) \\ &= \mathbf{f}\big(\mathbf{g}(\mathbf{a})\big) + d\mathbf{f}_{\mathbf{b}}\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w})\big) + d\mathbf{f}_{\mathbf{b}}\big(r_1(\mathbf{w})\big) \\ &\qquad + r_2\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big), \end{aligned} \tag{32}$$

having used the fact that $d\mathbf{f}_{\mathbf{b}}$ is a linear function.

Combining (31) and (32) gives

$$(\mathbf{f}\circ\mathbf{g})\left(\mathbf{a}+\mathbf{w}\right) = (\mathbf{f}\circ\mathbf{g})\left(\mathbf{a}\right) + \left(d\mathbf{f}_{\mathbf{b}}\circ d\mathbf{g}_{\mathbf{a}}\right)\left(\mathbf{w}\right) + d\mathbf{f}_{\mathbf{b}}\big(r_1(\mathbf{w})\big) + r_2\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big).$$

Substituting this into (30) gives

$$R(\mathbf{w}) = d\mathbf{f}_{\mathbf{b}}\big(r_1(\mathbf{w})\big) + r_2\big(d\mathbf{g}_{\mathbf{a}}\left(\mathbf{w}\right) + r_1(\mathbf{w})\big).$$

Thus, for $\mathbf{w} \neq \mathbf{0}$,

$$\frac{R(\mathbf{w})}{|\mathbf{w}|} = \frac{d\mathbf{f}_{\mathbf{b}}\big(r_1(\mathbf{w})\big)}{|\mathbf{w}|} + \frac{r_2\big(d\mathbf{g}_{\mathbf{a}}(\mathbf{w}) + r_1(\mathbf{w})\big)}{|\mathbf{w}|}. \tag{33}$$

2) *Proof of* $R(\mathbf{w})/|\mathbf{w}| \to \mathbf{0}$ as $\mathbf{w} \to 0$.

2i) *First term on RHS*(33) . Again using the fact that $d\mathbf{g_b}$ is linear, we see that the first term on the right satisfies

$$\frac{d\mathbf{f_b}\big(r_1(\mathbf{w})\big)}{|\mathbf{w}|} = d\mathbf{f_b}\left(\frac{r_1(\mathbf{w})}{|\mathbf{w}|}\right) \to d\mathbf{f_b}(\mathbf{0}) = \mathbf{0}, \tag{34}$$

as $\mathbf{w} \to \mathbf{0}$. This is because $r_1(\mathbf{w})/|\mathbf{w}| \to \mathbf{0}$ as $\mathbf{w} \to \mathbf{0}$ and $d\mathbf{f_b}$ is continuous at $\mathbf{0}$ (being a linear function).

2ii) *For the second term on RHS*(33) first consider $|d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})|$.

For the linear function $d\mathbf{g_a}$ there exists $C > 0$ such that $|d\mathbf{g_a}(\mathbf{w})| < C\,|\mathbf{w}|$ for all $\mathbf{w}$.

We are also assuming that $r_1(\mathbf{w})/|\mathbf{w}| \to \mathbf{0}$ as $\mathbf{w} \to \mathbf{0}$. Take $\varepsilon = 1$ in the definition of convergence to find $\delta_1 > 0$ such that if $|\mathbf{w}| < \delta_1$ then $|r_1(\mathbf{w})/|\mathbf{w}| - \mathbf{0}| < 1$, i.e. $|r_1(\mathbf{w})| \le |\mathbf{w}|$ (equality when $\mathbf{w} = \mathbf{0}$).

Combine to get that if $|\mathbf{w}| < \delta_1$ then, starting with the triangle inequality,

$$|d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})| \le |d\mathbf{g_a}(\mathbf{w})| + |r_1(\mathbf{w})| \le (C+1)\,|\mathbf{w}|\,. \tag{35}$$

Next consider $r_2\big(d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})\big)$. Let $\varepsilon > 0$ be given. Since $r_2(\mathbf{u})/|\mathbf{u}| \to \mathbf{0}$ as $\mathbf{u} \to \mathbf{0}$ there exists $\delta_2 > 0$ such that if $|\mathbf{u}| < \delta_2$ then $|r_2(\mathbf{u})/|\mathbf{u}|| < \varepsilon/(c+1)$, i.e.

$$|r_2(\mathbf{u})| < \frac{\varepsilon}{C+1}\,|\mathbf{u}|\,. \tag{36}$$

We will apply this with $\mathbf{u} = d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})$.

If we demand $(c+1)\,|\mathbf{w}| < \delta_2$ (along with $|\mathbf{w}| < \delta_1$) then (35) implies that

$$|d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})| < \delta_2$$

in which case

$$\left|r_2\big(d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})\big)\right| < \frac{\varepsilon}{C+1}\,|d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})|\,,$$

by (36). Use (35) again to get, for $|\mathbf{w}| < \min(\delta_1, \delta_2/(C+1))$,

$$\left|r_2\big(d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})\big)\right| < \frac{\varepsilon}{C+1}\,(C+1)\,|\mathbf{w}|\,, \quad \text{i.e.} \quad \frac{\left|r_2\big(d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})\big)\right|}{|\mathbf{w}|} < \varepsilon.$$

True for all $\varepsilon > 0$ implies

$$\lim_{\mathbf{w}\to\mathbf{0}} \frac{\left|r_2\big(d\mathbf{g_a}(\mathbf{w}) + r_1(\mathbf{w})\big)\right|}{|\mathbf{w}|} = \mathbf{0}.$$

Back in (33) with (34) gives us $\lim_{\mathbf{w}\to\mathbf{0}} R(\mathbf{w})/|\mathbf{w}| = \mathbf{0}$ as required. ■