DIFFERENTIABILITY

The chain rule.

Theorem: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}^\ell$, $x_0 \in \mathcal{U}$
$\mathcal{V} \subset \mathbb{R}^\ell$ open, $g : \mathcal{V} \longrightarrow \mathbb{R}^k$.

Assume that $f(\mathcal{U}) \subset \mathcal{V}$ so that $g \circ f : \mathcal{U} \longrightarrow \mathbb{R}^k$ is well-defined

If $\begin{cases} f \text{ is differentiable at } x_0 \\ g \text{ is differentiable at } f(x_0) \end{cases}$ then $\begin{cases} g\circ f \text{ is differentiable at } x_0 \text{ and} \\ d_{x_0}(g \circ f) = (d_{f(x_0)} g) \circ (d_{x_0} f) \end{cases}$

Δ $f(x_0 + h) = f(x_0) + d_{x_0}f(h) + \|h\|\, \varepsilon_1(h)$, $\varepsilon_1(h) \xrightarrow[h \to 0]{} 0$

$g(f(x_0) + k) = g(f(x_0)) + d_{f(x_0)}g(k) + \|k\|\, \varepsilon_2(k)$, $\varepsilon_2(k) \xrightarrow[k \to 0]{} 0$

Hence
$$\begin{aligned} g\circ f(x_0 + h) &= g(f(x_0+h)) \\ &= g(f(x_0) + d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)) \\ &= g(f(x_0)) + d_{f(x_0)}g(d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)) \\ &\quad + \|d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)\| \cdot \varepsilon_2(d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)) \\ &= g(f(x_0)) + d_{f(x_0)}g(d_{x_0}f(h)) \\ &\quad + \|h\| \Big(\underbrace{d_{f(x_0)}g(\varepsilon_1(h))}_{\xrightarrow[h\to 0]{} 0} + \underbrace{\|d_{x_0}f(\tfrac{h}{\|h\|}) + \varepsilon_1(h)\|}_{\text{bounded}} \cdot \underbrace{\varepsilon_2(\overbrace{d_{x_0}f(h) + \varepsilon_1(h)}^{\xrightarrow[h\to 0]{} 0})}_{\xrightarrow[h\to 0]{} 0 \text{ since } \varepsilon_2 \xrightarrow[h\to 0]{} 0} \Big) \end{aligned}$$

bounded:
- $\frac{h}{\|h\|} \in S^1$ compact and $d_{x_0}f$ continuous since linear
- $\varepsilon_1(h) \xrightarrow[h\to 0]{} 0$

$\square$

Corollary: Under the same assumptions

$$\underbrace{D(g\circ f)(x_0)}_{\substack{\text{Jacobian matrix of}\\ g\circ f \text{ at } x_0}} = \underbrace{Dg(f(x_0))}_{\substack{\text{Jacobian matrix}\\ \text{of } g \text{ at } f(x_0)}} \cdot \underbrace{Df(x_0)}_{\substack{\text{Jacobian matrix of } f\\ \text{at } x_0}}$$

Δ Recall that $Df(x_0) = \mathrm{Mat}(d_{x_0}f)$

and that $\mathrm{Mat}(d_{f(x_0)}g \circ d_{x_0}f) = \mathrm{Mat}(d_{f(x_0)}g)\,\mathrm{Mat}(d_{x_0}f)$ $\square$

Remark: If we look at the (i,j)-component, we get:

$$\frac{\partial(g_i\circ f)}{\partial x_j}(x_0) = \sum_{\alpha=1}^{\ell} \frac{\partial g_i}{\partial y_\alpha}(f(x_0)) \cdot \frac{\partial f_\alpha}{\partial x_j}(x_0)$$

[Recall that: $f: \mathbb{R}^m \longrightarrow \mathbb{R}^\ell$, $x \longmapsto f(x)$ and $g: \mathbb{R}^\ell \longrightarrow \mathbb{R}^k$, $y \longmapsto g(y)$]

Comment: Physicist notation "à la Leibniz":

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial v_i}{\partial y_1}\cdot\frac{\partial y_1}{\partial x_j} + \dots + \frac{\partial v_i}{\partial y_\ell}\cdot\frac{\partial y_\ell}{\partial x_j}$$

where $v = g(\overbrace{f(x)}^{y}) = (g_1(f(x)), \dots, g_k(f(x)))$, $y = f(x)$

$\frac{\partial v_i}{\partial x_j} = \frac{\partial (g_i\circ f)}{\partial x_j}$: we see $v(x) = g\circ f(x)$, as a function of $x \in \mathbb{R}^m$

$\frac{\partial v_i}{\partial y_\alpha} = \frac{\partial g_i}{\partial y_\alpha}$: we see $v(y) = g(y)$ as a function of $y \in \mathbb{R}^\ell$

$\partial y_\alpha/\partial x_j = \partial f_\alpha/\partial y_j$: $y(x) = f(x)$

Ex: Let $f: \mathbb{R}^2 \to \mathbb{R}$ differentiable

Define $\varphi: \mathbb{R}^3 \to \mathbb{R}$ by $\varphi(x,y,z) = f(x^2 - yz, xyz)$

$$\frac{\partial \varphi}{\partial x}(x,y,z) = 2x \frac{\partial f}{\partial x}(x^2-yz, xyz) + yz \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$$\frac{\partial \varphi}{\partial y}(x,y,z) = -z \frac{\partial f}{\partial x}(x^2-yz, xyz) + xz \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$$\frac{\partial \varphi}{\partial z}(x,y,z) = -y \frac{\partial f}{\partial x}(x^2-yz, xyz) + xy \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$\frac{\partial f}{\partial x}$ is just a notation to say partial derivative w.r.t. first variable

You compute the partial derivative of f wrt the first variable and then evaluate it at (x^2-yz, xyz)

Be sure you understand the above computations before continuing...

Your worst enemy in calculus is going to be notation:

① There are as many notation as people

eg: $\partial_x, \partial_{e_1}, \partial^{e_1}, \frac{\partial}{\partial x}, f_x, D^{(1,0)}, \ldots$ In MAT237, we'll use $\partial_x f$ or $\frac{\partial f}{\partial x}$.

Be sure that you understand what you are computing

② Notation can be confusing

eg: $\frac{\partial f}{\partial x}(x^2-yz, xyz)$ means ⓐ compute $\frac{\partial f}{\partial x}$ first partial derivative of f then ⓑ evaluate at (x^2-yz, xyz) → ✓

It does NOT mean: compute $f(x^2-yz, xyz)$ and then differentiate wrt to x → ✗

<u>Ex:</u> $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ differentiable

$S = \{(x,y) \in \mathbb{R}^2 : y \neq 0\}$ open

Define $\varphi(x,y) = f(x, xy, x/y)$

$$\frac{\partial \varphi}{\partial x}(x,y) = \frac{\partial f}{\partial x}(x, xy, x/y) + y\frac{\partial f}{\partial y}(x, xy, x/y) + \frac{1}{y}\frac{\partial f}{\partial z}(x, xy, x/y)$$

$$\frac{\partial \varphi}{\partial y}(x,y) = 0 + x\frac{\partial f}{\partial y}(x, xy, x/y) - \frac{x}{y^2}\frac{\partial f}{\partial z}(x, xy, x/y)$$

<u>Comment:</u>

$$D\varphi = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & x \\ 1/y & -\frac{x}{y^2} \end{pmatrix}$$

We recover the same result!

⚠ It's common to drop the variables during the computations as I did in the comment, in order to lighten the notation.

If you do so: ① Be careful to not forget to add them back at the end

② Keep track of them

Ex: Polar coordinates

$f: \mathbb{R}^2 \to \mathbb{R}$, $S = \{(r,\theta) \in \mathbb{R}^2 : r \geq 0\}$

$g: S \to \mathbb{R}^2$ defined by $g(r,\theta) = (r\cos\theta, r\sin\theta)$

$\varphi = f \circ g$ so that $\varphi(r,\theta) = f(r\cos\theta, r\sin\theta)$

Ⓐ Chain rule for the Jacobian matrix:

$$\begin{aligned} D\varphi(r,\theta) &= Df(g(r,\theta)) \cdot Dg(r,\theta) \\ &= Df(r\cos\theta, r\sin\theta) \cdot Dg(r,\theta) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) & \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta) \end{pmatrix} \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \end{aligned}$$

$$\begin{pmatrix} \frac{\partial \varphi}{\partial r}(r,\theta) & \frac{\partial \varphi}{\partial \theta}(r,\theta) \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\cos\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\sin\theta , -\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta \end{pmatrix}$$

Ⓑ Chain rule for the partial derivatives

$$\frac{\partial \varphi}{\partial r}(r,\theta) = \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\cos\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\sin\theta$$

$$\frac{\partial \varphi}{\partial \theta}(r,\theta) = -\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\, r\sin\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\, r\cos\theta$$

We read them in the components of the Jacobian matrix

Ⓒ Chain rule for the differentials:

$$\begin{aligned} d_{(r,\theta)}\varphi(h,k) &= d_{g(r,\theta)}f \circ d_{(r,\theta)}g(h,k) = d_{g(r,\theta)}f(\cos\theta h - r\sin\theta k, \sin\theta h + r\cos\theta k) \\ &= \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)(\cos\theta h - r\sin\theta k) + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)(\sin\theta h + r\cos\theta k) \end{aligned}$$