Topic(s): multivariate chain rule

Example 1. Chris Sharma, a world-class rock climber, is training for the upcoming Tokyo Olympics. The mountain he has chosen to climb today has height h given by $h(x, y) = 200 - 3x^2 - xy - y^2$. Chris' position at time t is given by $\mathbf{p}(t) = \begin{bmatrix} t-20 \\ 3t-60 \end{bmatrix}$. Compute the rate of change of Chris' height at time t. $= (h \circ \vec{p})'(t)$

We see that

$$(h \circ \vec{p})(t) = h(t-20, 3t-60)$$
$$= 200 - 3(t-20)^2 - (t-20)(3t-60) - (3t-60)^2$$
$$= -15t^2 + 600t - 5800.$$

Hence, $\boxed{(h \circ \vec{p})'(t) = -30t + 600}$.

Example 2. Suppose $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $\mathbf{F} = \mathbf{f} \circ \mathbf{g} \circ \mathbf{h}$, where $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$, and $\mathbf{h} : \mathbb{R}^2 \to \mathbb{R}^3$ are defined by

$$\mathbf{f}(v, w) = \begin{bmatrix} vw \\ v+w \end{bmatrix} \qquad \mathbf{g}(x, y, z) = \begin{bmatrix} xy \\ 3x + yz \end{bmatrix} \qquad \mathbf{h}(s, t) = \begin{bmatrix} s^2 t \\ st^2 \\ s+t \end{bmatrix}.$$

Compute $\left.\dfrac{\partial F_2}{\partial t}\right|_{(2,-1)}$.

We first need to compute F_2 in terms of s and t.

$$F_2(s,t) = f_2 \circ \vec{g} \circ \vec{h}(s,t)$$
$$= f_2 \circ \vec{g}(s^2t, st^2, s+t)$$
$$= f_2((s^2t)(st^2), 3(s^2t) + (st^2)(s+t))$$
$$= f_2(s^3t^3, 3s^2t + s^2t^2 + st^3)$$
$$= s^3t^3 + 3s^2t + s^2t^2 + st^3.$$

Hence, $\dfrac{\partial F_2}{\partial t}(s,t) = 3s^3t^2 + 3s^2 + 2s^2t + 3st^2$, and so, $\left.\dfrac{\partial F_2}{\partial t}\right|_{(2,-1)} = \boxed{34}$.

Theorem 17.1.5 (Chain rule). If $\mathbf{f} : \mathbb{R}^p \to \mathbb{R}^m$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^p$, then the derivative matrix of $\mathbf{f} \circ \mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ at $\mathbf{a} \in \mathbb{R}^n$ is

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{g}(\mathbf{a}))\; D\mathbf{g}(\mathbf{a}).$$

Note that the right-hand side is a *matrix product.* Intuitively speaking, this is equivalent to saying "the linear approximation of $\mathbf{f} \circ \mathbf{g}$ at $\mathbf{a} \in \mathbb{R}^n$ is the same as first linearly approximating $\mathbf{g}$ at $\mathbf{a}$, then linear approximating $\mathbf{g}$ at your result (in $\mathbb{R}^p$). Similarly, if $\mathbf{x} \in \mathbb{R}^n$ is a variable vector, then

$$\underbrace{D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})}_{m \times n} = \underbrace{D\mathbf{f}(\mathbf{g}(\mathbf{x}))}_{m \times p}\; \underbrace{D\mathbf{g}(\mathbf{x})}_{p \times n}.$$

$$\frac{\partial f}{\partial x_j} = \sum_{k=1}^{p} \frac{\partial f}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j} \quad (17.1.6)$$

is the version of the chain rule in most calculus books.

Example 2.5. Suppose $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $\mathbf{F} = \mathbf{f} \circ \mathbf{g} \circ \mathbf{h}$, where $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$, and $\mathbf{h} : \mathbb{R}^2 \to \mathbb{R}^3$ are defined by

$$\mathbf{f}(v,w) = \begin{bmatrix} vw \\ v+w \end{bmatrix} \qquad \mathbf{g}(x,y,z) = \begin{bmatrix} xy \\ 3x+yz \end{bmatrix} \qquad \mathbf{h}(s,t) = \begin{bmatrix} s^2t \\ st^2 \\ s+t \end{bmatrix}.$$

Compute $\left.\dfrac{\partial F_2}{\partial t}\right|_{(2,-1)}$.

We see that $\vec{h}(2,-1) = (-4,2,1)$ and $\vec{g}(-4,2,1) = (-8,-10)$. The derivative matrices are

$$D\vec{f} = \begin{bmatrix} w & v \\ 1 & 1 \end{bmatrix}, \qquad D\vec{g} = \begin{bmatrix} y & x & 0 \\ 3 & z & y \end{bmatrix}, \qquad \text{and} \qquad D\vec{h} = \begin{bmatrix} 2st & s^2 \\ t^2 & 2st \\ 1 & 1 \end{bmatrix}.$$

So, $D\vec{f}(-8,-10) = \begin{bmatrix} -10 & -8 \\ 1 & 1 \end{bmatrix}$, $D\vec{g}(-4,2,1) = \begin{bmatrix} 2 & -4 & 0 \\ 3 & 1 & 2 \end{bmatrix}$, and $D\vec{h}(2,-1) = \begin{bmatrix} -4 & 4 \\ 1 & -4 \\ 1 & 1 \end{bmatrix}$.

By the chain rule,

$$D\vec{F}(2,-1) = \begin{bmatrix} -10 & -8 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -4 & 0 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} -4 & 4 \\ 1 & -4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 192 & -320 \\ -21 & 34 \end{bmatrix}.$$

$(F_2)_t(2,-1)$

Example 3. Suppose the temperature in a room is given by $T(x,y,z) = x^2 - 4x + y^2 + e^z$. A ladybug begins at rest on the floor at $(4,2,0)$ flies around along a spiral path $\mathbf{p}(t) = (3+\cos t, 2+\sin t, t)$, where t is the time parameter. At $t=3$, what is the rate of change with respect to time for the temperature experienced by the ladybug along its path of motion?

$= (T \circ \vec{p})'(3)$

We get $\vec{P}(3) = (3+\cos 3, 2+\sin 3, 3)$, $DT = [2x-4 \quad 2y \quad e^z]$, and $D\vec{p} = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix}$.

So, $DT(\vec{p}(3)) = [2+2\cos 3 \quad 4+2\sin 3 \quad e^3]$ and $D\vec{p}(3) = \begin{bmatrix} -\sin 3 \\ \cos 3 \\ 1 \end{bmatrix}$. By the chain rule,

$$D(T\circ\vec{p})(3) = [2+2\cos 3 \quad 4+2\sin 3 \quad e^3] \begin{bmatrix} -\sin 3 \\ \cos 3 \\ 1 \end{bmatrix} = [-2\sin 3 + 4\cos 3 + e^3]$$

Thus, $(T\circ\vec{p})'(3) = -2\sin 3 + 4\cos 3 + e^3$.

To evaluate a composite derivative matrix at $\mathbf{a}$, we have two options:

- **Numerical method.** Compute $D\mathbf{f}(\mathbf{g}(\mathbf{a}))$ and $D\mathbf{g}(\mathbf{a})$, then multiply them.
- **Symbolic method.** Compute $D(\mathbf{f}\circ\mathbf{g})$ in general, then evaluate at $\mathbf{a}$.

Usually the numerical method is simpler, but the symbolic method can be useful if we have a lot of points to compute.

Example 4. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ and $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$\mathbf{f}(x,y,z) = \begin{bmatrix} x^2+y^2+z^2 \\ xz+y^2 \end{bmatrix} \qquad \mathbf{g}(s,t) = \begin{bmatrix} t^2 \\ st \\ \frac{1}{s} \end{bmatrix}$$

Evaluate $D(\mathbf{f}\circ\mathbf{g})(4,2)$ by using

- the symbolic method.

Since $D\vec{f} = \begin{bmatrix} 2x & 2y & 2z \\ z & 2y & x \end{bmatrix}$ and $D\vec{g} = \begin{bmatrix} 0 & 2t \\ t & s \\ -\frac{1}{s^2} & 0 \end{bmatrix}$, we have

$$D(\vec{f}\circ\vec{g})(s,t) = \begin{bmatrix} 2t^2 & 2st & \frac{2}{s} \\ \frac{1}{s} & 2st & t^2 \end{bmatrix}\begin{bmatrix} 0 & 2t \\ t & s \\ -\frac{1}{s^2} & 0 \end{bmatrix} = \begin{bmatrix} 2st^2 - \frac{2}{s^3} & 4t^3 + 2s^2t \\ 2st^2 - \frac{t^2}{s^2} & \frac{2t}{s} + 2s^2t \end{bmatrix}$$

by the chain rule. Hence, $D(\vec{f}\circ\vec{g})(4,2) = \boxed{\begin{bmatrix} \frac{1023}{32} & 96 \\ \frac{127}{4} & 65 \end{bmatrix}}$.

- the numerical method.

We see that $\vec{g}(4,2) = (4, 8, \frac{1}{4})$,

$$D\vec{g}(4,2) = \begin{bmatrix} 0 & 4 \\ 2 & 4 \\ -\frac{1}{16} & 0 \end{bmatrix}, \quad \text{and} \quad D\vec{f}(4,8,\tfrac{1}{4}) = \begin{bmatrix} 8 & 16 & \frac{1}{2} \\ \frac{1}{4} & 16 & 4 \end{bmatrix}.$$

Hence, by the chain rule, $D(\vec{f}\circ\vec{g}) = \begin{bmatrix} 8 & 16 & \frac{1}{2} \\ \frac{1}{4} & 16 & 4 \end{bmatrix}\begin{bmatrix} 0 & 4 \\ 2 & 4 \\ -\frac{1}{16} & 0 \end{bmatrix} = \boxed{\begin{bmatrix} \frac{1023}{32} & 96 \\ \frac{127}{4} & 65 \end{bmatrix}}$.

Example 5. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$\mathbf{f}(x, y) = \begin{bmatrix} x^2 + 3xy - y^2 - y + 1 \\ 2x^2 - xy + y^2 + 3x - 4 \end{bmatrix}.$$

(a) Compute $(D\mathbf{f})(0, 0)$.

We compute

$$D\vec{f} = \begin{bmatrix} 2x+3y & 3x-2y-1 \\ 4x-y+3 & -x+2y \end{bmatrix},$$

and so, $D\vec{f}(0,0) = \begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}$.

(b) Suppose $\mathbf{h} : \mathbb{R} \to \mathbb{R}^2$ with $\mathbf{h}(0) = (0, 0)$ and $(D\mathbf{h})(0) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Use linear approximations to estimate $(\mathbf{f} \circ \mathbf{h})(0.1)$.

We compute $(\vec{f} \circ \vec{h})(0) = \vec{f}(0,0) = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, and so, by chain rule,

$$D(\vec{f} \circ \vec{h})(0) = D\vec{f}(0,0)\, D\vec{h}(0) = \begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}.$$

Hence,

$$(\vec{f} \circ \vec{h})(0.1) \approx (\vec{f} \circ \vec{h})(0) + D(\vec{f} \circ \vec{h})(0)\,(0.1 - 0)$$
$$= \begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 \\ 6 \end{bmatrix}[0.1] = \boxed{\begin{bmatrix} 1.1 \\ -3.4 \end{bmatrix}}.$$

(c) Let $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $\mathbf{g} = \mathbf{f} \circ \mathbf{f}$. Compute $(D\mathbf{g})(0, 0)$.

By chain rule,

$$(D\vec{g})(0,0) = (D\vec{f})(1,-4)(D\vec{f})(0,0) = \begin{bmatrix} -10 & 10 \\ 11 & -9 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}$$
$$= \boxed{\begin{bmatrix} 30 & 10 \\ -27 & -11 \end{bmatrix}}.$$

Exercise 17.8 (a). Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\mathbf{v} \in \mathbb{R}^n$ be a unit vector. The **directional derivative** of f in direction of $\mathbf{v}$ at a point $\mathbf{a} \in \mathbb{R}^n$ is defined to be

$$(D_{\mathbf{v}} f)(\mathbf{a}) = \left.\frac{d}{dt}\right|_{t=0} f(\mathbf{a} + t\mathbf{v}) \in \mathbb{R}$$

(the rate of change of f at times $t = 0$ when viewed as a function on the line through $\mathbf{a}$ in the direction of $\mathbf{v}$ that is parametrized at unit speed in that direction). By writing $f(\mathbf{a} + t\mathbf{v})$ as $(f \circ \mathbf{g})(t)$ for $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ defined by $\mathbf{g}(t) = \mathbf{a} + t\mathbf{v}$, use the Chain Rule to show that $(D_{\mathbf{v}} f)(\mathbf{a}) = ((Df)(\mathbf{a}))\mathbf{v}$ (a matrix-vector product).

Example 6. We continue the discussion of directional derivatives. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $\mathbf{v} \in \mathbb{R}^n$ a unit vector and $\mathbf{a} \in \mathbb{R}^n$ a point.

(a) Show that $D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}$.

By the above result,

$$\begin{aligned} D_{\vec{v}} f(\vec{a}) &= ((Df)(\vec{a}))\,\vec{v} \\ &= \begin{bmatrix} f_{x_1}(\vec{a}) & f_{x_2}(\vec{a}) & \cdots & f_{x_n}(\vec{a}) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= f_{x_1}(\vec{a})v_1 + f_{x_2}(\vec{a})v_2 + \cdots + f_{x_n}(\vec{a})v_n \\ &= \nabla f(\vec{a}) \cdot \vec{v}. \end{aligned}$$

(b) Assuming $\nabla f(\mathbf{a})$ is non-zero, in what unit direction is the directional derivative of f the biggest? That is, for which unit vector(s) $\mathbf{v}$ is $D_{\mathbf{v}} f(\mathbf{a})$ the biggest?

Since $\vec{v}$ is a unit vector,

$$\begin{aligned} D_{\vec{v}} f(\vec{a}) &= \nabla f(\vec{a}) \cdot \vec{v} \\ &= \|\nabla f(\vec{a})\| \, \|\vec{v}\| \cos\theta \quad \leftarrow \text{angle between } \nabla f(\vec{a}) \text{ and } \vec{v} \\ &= \|\nabla f(\vec{a})\| \cos\theta. \end{aligned}$$

$\|\nabla f(\vec{a})\|$ is fixed, and so, $D_{\vec{v}} f(\vec{a})$ depends only on θ. $\cos\theta$ is maximized at $\theta = 0°$.
($\vec{v}$ is in the same direction as $\nabla f(\vec{a})$, i.e., $\nabla f(\vec{a})$ is the direction of fastest increase.)
Similarly, $\cos\theta$ is minimized at $\theta = 180°$ ($-\nabla f(\vec{a})$ is the direction of fastest decrease)

Example 7. Let $\mathbf{f}, \mathbf{g} : \mathbb{R} \to \mathbb{R}^2$ be two vector-valued functions. Use the chain rule to compute that for the function h defined by the dot product $h(x) = \mathbf{f}(x) \cdot \mathbf{g}(x)$, the 1×1 derivative matrix $Dh(x)$ is given by the "product rule" formula

$$Dh(x) = \mathbf{g}(x)^\intercal D\mathbf{f}(x) + \mathbf{f}(x)^\intercal D\mathbf{g}(x),$$

where the notation $\mathbf{w}^\intercal$ for an m-vector $\mathbf{w}$ is the $1 \times m$ row vector with w_i as its ith entry.

We can think of the dot product function as the composition concatenating $\vec{f}(x)$ and $\vec{g}(x)$ into one 4-vector and a function that computes the dot product from the 4-vector. More specifically, $h = \psi \circ \vec{\varphi}$, where

$$\vec{\varphi}(x) = (\vec{f}(x), \vec{g}(x)) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ g_1(x) \\ g_2(x) \end{bmatrix} \quad \text{and} \quad \psi(x_1, \dots, x_4) = x_1 x_3 + x_2 x_4.$$

The derivative matrices are

$$D\vec{\varphi} = \begin{bmatrix} f_1'(x) \\ f_2'(x) \\ g_1'(x) \\ g_2'(x) \end{bmatrix} = \begin{bmatrix} D\vec{f} \\ D\vec{g} \end{bmatrix} \quad \text{and} \quad D\psi = \begin{array}{c} \begin{bmatrix} x_3 & x_4 & x_1 & x_2 \end{bmatrix} \\ \;\; g_1 \;\;\; g_2 \;\;\; f_1 \;\;\; f_2 \end{array} = [\vec{g}(x)^\intercal \;\; \vec{f}(x)^\intercal].$$

Therefore,

$$Dh = D\psi D\vec{\varphi} = [\vec{g}(x)^\intercal \;\; \vec{f}(x)^\intercal] \begin{bmatrix} D\vec{f} \\ D\vec{g} \end{bmatrix} = \vec{g}(x)^\intercal D\vec{f} + \vec{f}(x)^\intercal D\vec{g}.$$