Lecture 17 - Multivariable Chain Rule

November 4, 2022

Goals: Translate word problems involving composite functions into a mathematical statement, use the multivariable chain rule to find the derivative matrix of composite functions.

Example 1: A person wants to test their rock climbing skills on an actual mountain. Suppose this mountain has height h given by $h(x,y) = 200 - 3x^2 - xy - y^2$. Their position at time t is $\mathbf{p}(t) = \begin{bmatrix} t-10 \\ 2t-30 \end{bmatrix}$. Compute the rate of change of the person's height at time t.

$\left[\; p(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}\right]$

$$\frac{d}{dt} h(x,y) = \frac{d}{dt}(h \circ \vec{p})(t)$$

$$= \frac{d}{dt}\left(200 - 3(t-10)^2 - (t-10)(2t-30) - (2t-30)^2\right) = \frac{\partial h}{\partial x}\left(\frac{dx}{dt}\right) + \frac{\partial h}{\partial y}\left(\frac{dy}{dt}\right)$$

$$= -6(t-10) - 2(t-10) - (2t-30) - 2(2t-30)2$$

$$= \boxed{-18t + 230}$$

$$= (-6x - y)(1) + (-x - 2y)(2)$$

$$= (-6(t-10) - (2t-30)) + (-(t-10) - 2(2t-30))(2)$$

$$= -6t + 60 - 2t + 30 - 2t + 20 - 8t + 120$$

$$= -18t + 230$$

Example 2: Suppose $F : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $F = f \circ g \circ h$ where each function is defined via

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

What is $\frac{\partial F_2}{\partial t}(2,-1)$? Here, F_2 is the second component of F.

$$F_2(s,t) = f_2 \circ g \circ h(s,t)$$

$$f_2(g(h(s,t))) = f_2(g(s^2t, st^2, s+t))$$

$$= f_2(s^3t^3, 2s^2t + s^2t^2 + st^3)$$

$$= s^3t^3 + 2s^2t + s^2t^2 + st^3$$

$$\frac{\partial F_2}{\partial t} = 3s^3t^2 + 2s^2 + 2s^2t + 3st^2$$

$$\frac{\partial F_2}{\partial t}(2,-1) = 30$$

$$f \circ g : \mathbb{R}^n \to \mathbb{R}^p \to \mathbb{R}^m$$

Theorem 17.1.5: If $f : \mathbb{R}^p \to \mathbb{R}^m$, and $g : \mathbb{R}^n \to \mathbb{R}^p$ are two functions then the derivative matrix of $f \circ g : \mathbb{R}^n \to \mathbb{R}^m$ at a point $\mathbf{a} \in \mathbb{R}^n$ is

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

Note that this mimics the form of the single variable chain rule: $(f(g(a)))' = f'(g(a))g'(a)$. Indeed, the single variable chain rule is just a special case, since the derivative matrix of a single variable $f(x)$ is exactly the 1×1 matrix $f'(x)$.

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}, \quad g = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \quad \text{implies} \quad (i,j) \text{ entry of } D(f \circ g) = \frac{\partial f_i}{\partial x_j} = \sum_{k=1}^{p} \frac{\partial f_i}{\partial y_k}\frac{\partial y_k}{\partial x_j}$$

Example 3: Redo Example 2 by using the multivariable chain rule as stated above. Recall that we were asked to find $\frac{\partial F_2}{\partial t}(2,-1)$, with $F = f \circ g \circ h$, where

$$DF = \begin{bmatrix} \frac{\partial F_1}{\partial s} & \frac{\partial F_1}{\partial t} \\ \frac{\partial F_2}{\partial s} & \frac{\partial F_2}{\partial t} \end{bmatrix}$$

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

$$D(F)(2,-1) = \underbrace{D(f \circ g \circ h)(2,-1)}_{2\times2} = \underbrace{Df(g(h(2,-1)))}_{2\times2} \cdot \underbrace{Dg(h(2,-1))}_{2\times3} \cdot \underbrace{Dh(2,-1)}_{3\times2}$$

$$h(2,-1) = (-4, 2, 1), \quad g(h(2,-1)) = g(-4,2,1) = (-8,-6)$$

$$= \underbrace{Df(-8,-6)}_{2\times2} \cdot \underbrace{Dg(-4,2,1)}_{2\times3} \cdot \underbrace{Dh(2,-1)}_{3\times2}$$

$$Df = \begin{bmatrix} w & v \\ 1 & 1 \end{bmatrix} \quad ; \quad Df(-8,-6) = \begin{bmatrix} -6 & -8 \\ 1 & 1 \end{bmatrix}$$

$$Dg = \begin{bmatrix} y & x & 0 \\ 2 & z & y \end{bmatrix} \quad ; \quad Dg(-4,2,1) = \begin{bmatrix} 2 & -4 & 0 \\ 2 & 1 & 2 \end{bmatrix}$$

$$Dh = \begin{bmatrix} 2st & s^2 \\ t^2 & 2st \\ 1 & 1 \end{bmatrix} \quad ; \quad Dh(2,-1) = \begin{bmatrix} -4 & 4 \\ 1 & -4 \\ 1 & 1 \end{bmatrix}$$

answer: $$D(F)(2,-1) = \begin{bmatrix} -6 & -8 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -4 & 0 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} -4 & 4 \\ 1 & -4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 136 & -240 \\ -17 & 30 \end{bmatrix}$$

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

$T: \mathbb{R}^3 \to \mathbb{R}$ $\qquad$ $T \circ \vec{p} \to$ derivative at $t=3$

$p: \mathbb{R} \to \mathbb{R}^3$

$$\underbrace{D(T \circ \vec{p})(3)}_{1\times1} = \underbrace{DT(\vec{p}(3))}_{1\times3} \cdot \underbrace{D\vec{p}(3)}_{3\times1}$$

$$DT(x,y,z) = \begin{bmatrix} 2x-4 & 2y & e^z \end{bmatrix}$$

$$D\vec{p}(t) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix} \to D\vec{p}(3) = \begin{bmatrix} -\sin 3 \\ \cos 3 \\ 1 \end{bmatrix}$$

$$DT(\vec{p}(3)) = \begin{bmatrix} 2(3+\cos 3) - 4 & 2(2+\sin 3) & e^3 \end{bmatrix}$$
$$= \begin{bmatrix} 2 + 2\cos 3 & 4 + 2\sin 3 & e^3 \end{bmatrix}$$

$$DT(\vec{p}(3)) \cdot D\vec{p}(3) = \begin{bmatrix} 2 + 2\cos 3 & 4 + 2\sin 3 & e^3 \end{bmatrix} \begin{bmatrix} -\sin 3 \\ \cos 3 \\ 1 \end{bmatrix}$$

$$= -2\sin 3 - 2\sin 3\cos 3 + 4\cos 3 + 2\sin 3\cos 3 + e^3$$
$$= -2\sin 3 + 4\cos 3 + e^3$$
$$\approx 15.84$$

In the previous examples, we have been computing the derivative matrices using the "**numerical**" method: computing each derivative matrix separately, plugging in the corresponding points, and then doing matrix multiplication. We could also use the "**symbolic**" method: Compute the derivative matrix of the composition as a function of $\mathbf{x}$, and then evaluate at $\mathbf{a}$. This method can be useful if we need to compute the derivative at a lot of points.

Example 5: Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$f(x,y,z) = \begin{bmatrix} x^2+y^2+z^2 \\ xy+z^2 \end{bmatrix}, \qquad g(s,t) = \begin{bmatrix} t^2 \\ st \\ \sqrt{s} \end{bmatrix}.$$

Evaluate $D(f\circ g)(2,2)$ using both the symbolic and numerical methods.

Symbolic:

$$D(f\circ g)(s,t) = Df(g(s,t)) \cdot Dg(s,t)$$

$$Df(x,y,z) = \begin{bmatrix} 2x & 2y & 2z \\ y & x & 2z \end{bmatrix}$$

$$Df(g(s,t)) = Df(t^2, st, \sqrt{s}) = \begin{bmatrix} 2t^2 & 2st & 2\sqrt{s} \\ st & t^2 & 2\sqrt{s} \end{bmatrix}$$

$$Dg(s,t) = \begin{bmatrix} 0 & 2t \\ t & s \\ \frac{1}{2\sqrt{s}} & 0 \end{bmatrix}$$

$$D(f\circ g)(s,t) = Df(g(s,t)) \cdot Dg(s,t) = \begin{bmatrix} 2t^2 & 2st & 2\sqrt{s} \\ st & t^2 & 2\sqrt{s} \end{bmatrix} \begin{bmatrix} 0 & 2t \\ t & s \\ \frac{1}{2\sqrt{s}} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2st^2+1 & 4t^3+2s^2t \\ t^3+1 & 2st^2+st^2 \end{bmatrix}$$

$$D(f\circ g)(2,2) = \begin{bmatrix} 17 & 48 \\ 9 & 24 \end{bmatrix}$$

Numerical

$$Df(g(s,t)) = Df(t^2, st, \sqrt{s}) = \begin{bmatrix} 2t^2 & 2st & 2\sqrt{s} \\ st & t^2 & 2\sqrt{s} \end{bmatrix}, \quad Df(4,4,\sqrt{2}) = \begin{bmatrix} 8 & 8 & 2\sqrt{2} \\ 4 & 4 & 2\sqrt{2} \end{bmatrix}$$

$$Dg(s,t) = \begin{bmatrix} 0 & 2t \\ t & s \\ \frac{1}{2\sqrt{s}} & 0 \end{bmatrix}, \quad Dg(2,2) = \begin{bmatrix} 0 & 4 \\ 2 & 2 \\ \frac{1}{2\sqrt{2}} & 0 \end{bmatrix}$$

$$D(f\circ g)(2,2) = \begin{bmatrix} 17 & 48 \\ 9 & 24 \end{bmatrix}$$

Example 6 (if time): Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

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

Suppose $h : \mathbb{R} \to \mathbb{R}^2$ is defined by

$$h(x) = \begin{bmatrix} 2x \\ -x \end{bmatrix}.$$

Use linear approximation to estimate $(f \circ h)(0.1)$.

$$f(\vec{x}) = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a})$$

$$(f \circ h)(\vec{x}) = (f \circ h)(\vec{a}) + D(f \circ h)(\vec{a})(\vec{x} - \vec{a})$$
$$= f(h(\vec{a})) + Df(h(\vec{a})) \cdot Dh(\vec{a})(\vec{x} - \vec{a})$$

$$(f \circ h)(0.1) = f(h(0)) + Df(h(0)) \cdot Dh(0)(0.1 - 0)$$
$$= f\begin{pmatrix} 0 \\ 0 \end{pmatrix} + Df\begin{pmatrix} 0 \\ 0 \end{pmatrix} \cdot Dh(0)(0.1 - 0)$$
$$= \begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix}(0.1)$$
$$= \begin{bmatrix} 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 \\ 6 \end{bmatrix}(0.1)$$
$$= \begin{bmatrix} 1.1 \\ -3.4 \end{bmatrix}$$

actual ans: $\begin{bmatrix} 0.71 \\ -3.29 \end{bmatrix}$

Suppose $f(g(x), h(x))$

When $x \uparrow$, you add up contribution of function g as a result of x $\Big\}$ rate $g \uparrow$ as $x \uparrow$ times rate $f \uparrow$ as $g \uparrow$ $\Big\}$ $g'(x) \cdot \frac{\partial}{\partial x_1} f(g(x), h(x))$

and of function h as a result of x $\Big\}$ rate $h \uparrow$ as $x \uparrow$ times rate $f \uparrow$ as $h \uparrow$ $\Big\}$ $h'(x) \cdot \frac{\partial}{\partial x_2} f(g(x), h(x))$

Consider $\mathbf{g} : \mathbf{R}^n \to \mathbf{R}^p$, and write its components as $g_1(\mathbf{x}), \ldots, g_p(\mathbf{x})$. For any $f : \mathbf{R}^p \to \mathbf{R}$, the Chain Rule formula for $f \circ \mathbf{g}$ can be written without reference to matrices as

$$\frac{\partial(f \circ \mathbf{g})}{\partial x_j} = \sum_{k=1}^{p} \frac{\partial f}{\partial y_k} \frac{\partial g_k}{\partial x_j}; \tag{17.1.5}$$

with $1 \leq j \leq n$; each $\partial f / \partial y_k$ on the right side is evaluated at $\mathbf{g}(\mathbf{x})$, and the left side is evaluated at $\mathbf{x}$.

Where does the right side of (17.1.5) come from? It is the expression for the $1j$-entry in the $1 \times n$ product matrix $(Df)(\mathbf{g}(\mathbf{x}))\,(D\mathbf{g})(\mathbf{x})$ (we say more about this in Remark 17.4.1). A convenient way to think about (17.1.5) is that by writing $\mathbf{y} = \mathbf{g}(\mathbf{x})$, it expresses rates of change of f in terms of the x's as a sum of contributions of rates of change of f in terms of the y's multiplied by rates of change of the y's in terms of the x's: abbreviating the notation in (17.1.5) by writing f instead of $f \circ \mathbf{g}$ on the left side and writing y_k instead of g_k on the right side expresses it as:

$$\frac{\partial f}{\partial x_j} = \sum_{k=1}^{p} \frac{\partial f}{\partial y_k} \frac{\partial y_k}{\partial x_j} \tag{17.1.6}$$

(where the left side is evaluated at $\mathbf{x}$ and each $\partial f / \partial y_k$ on the right side is evaluated at $\mathbf{g}(\mathbf{x})$). Although (17.1.5) is the more precise formulation, **you will often encounter the notationally convenient version** (17.1.6) **in many places.** Observe that when $n = 1$ this recovers (17.1.4).

Ex:

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

$$= \sum_{k=1}^{2} \frac{\partial f}{\partial x_k} \frac{\partial x_k}{\partial r}$$

$$= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$

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

$$\boxed{\frac{\partial h}{\partial r} = \cos\theta \frac{\partial f}{\partial x} + \sin\theta \frac{\partial f}{\partial y}}$$

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

$$= \sum_{k=1}^{2} \frac{\partial f}{\partial x_k} \frac{\partial x_k}{\partial \theta}$$

$$= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}$$

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

$$\boxed{\frac{\partial h}{\partial \theta} = -r\sin\theta \frac{\partial f}{\partial x} + r\cos\theta \frac{\partial f}{\partial y}}$$

$$D(f \circ g \circ h)(x) = Df((g \circ h)(x)) \cdot Dg(h(x)) \cdot Dh(x)$$

Ex: $f(x,y) = xy \qquad g(t) = (t^2, \sin(t))$

$$D(f \circ g)(t) = Df(g(t)) \cdot Dg(t)$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x}(g(t)) & \frac{\partial f}{\partial y}(g(t)) \end{bmatrix} \begin{bmatrix} g_1'(t) \\ g_2'(t) \end{bmatrix}$$

$$= \begin{bmatrix} \sin(t) & t^2 \end{bmatrix} \begin{bmatrix} 2t \\ \cos t \end{bmatrix}$$

$$= 2t\sin(t) + t^2\cos(t)$$

$$D(f \circ g)(t) = Df(g(t)) \cdot Dg(t) = \begin{bmatrix} \sin t & t^2 \end{bmatrix} \begin{bmatrix} 2t \\ \cos t \end{bmatrix}$$

$$Df(x,y) = \begin{bmatrix} y & x \end{bmatrix}$$

$$Df(g(t)) = Df(t^2, \sin t) = \begin{bmatrix} \sin t & t^2 \end{bmatrix}$$

$$Dg(t) = \begin{bmatrix} 2t \\ \cos t \end{bmatrix}$$