Chapter 14.4: The Chain Rule

D1: Computing Derivatives. I can compute partial derivatives, total derivatives, directional derivatives, and gradients. I can use the Chain Rule for multivariable functions to compute derivatives of composite functions.

A1: Interpreting Derivatives. I can interpret the meaning of a partial derivative, a gradient, or a directional derivative of a function at a given point in a specified direction, including in the context of a graph or a contour plot.

Mechanics

1. Use the chain rule to compute the total derivatives of the following at the prescribed points. [*Recall that Df for $f : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix.*]

   (a) $f : \mathbb{R} \to \mathbb{R}$ given by $f(t) = h(g(t))$, where $g(t) = (t+1, t^2)$, $h(x,y) = xy$ at $t = 2$.

   (b) $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = (b \circ a)(x,y)$, where $a(x,y) = x \sin y$, $b(t) = 8t - t^2$ at $(x,y) = (4, \pi/3)$.

   (c) $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = u(v(t))$, where $v(t) = (1, t, t^2)$, $u(x,y,z) = (xy, yz)$ at $t = 1$.

   (d) $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = h(g(x,y))$, where $g(x,y) = (3x + 4y, 5x + 7y)$, $h(u,v) = (7u - 4v, -5u + 3v)$ at $(x,y) = (0,0)$.

   What is interesting about (d)? How are g and h related?

   Answer:

   (a) $Df(2) = \begin{bmatrix} 16 \end{bmatrix}$

   (b) $Df(4, \pi/3) = \begin{bmatrix} 4\sqrt{3} - 6 & 16 - 8\sqrt{3} \end{bmatrix}$

   (c) $Df(1) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

   (d) $Df(0,0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

2. Find the values of t where $\dfrac{dz}{dt} = 0$ if $z = 3x + 4y$, $x = t^2$, and $y = 2t$.

   Answer: $t = -4/3$

3. Let $w(x,y,z) = xy + yz + zx$, where $x = r\cos\theta$, $y = r\sin\theta$, $z = r\theta$. Find $\dfrac{\partial w}{\partial r}$ and $\dfrac{\partial w}{\partial \theta}$ when $r = 2$ and $\theta = \dfrac{\pi}{2}$.

   Answer: $\dfrac{\partial w}{\partial r}(2, \pi/2) = 2\pi$ $\dfrac{\partial w}{\partial \theta}(2, \pi/2) = -2\pi$

Applications

4. You are a myrmecologist (ant scientist) studying the behaviour of ants on anthills. You equipped an ant with a tracker to track its motion. Unfortunately, the exact geometry of the hill, as well as the ant's precise motion are too delicate to measure. Fortunately, at time $t = 5$ seconds it is known that

$$\frac{\partial z}{\partial x} = 5, \quad \frac{\partial z}{\partial y} = -2, \quad \frac{dx}{dt} = 3, \quad \frac{dy}{dt} = 7.$$

where z denotes the height of the ant relative to ground height. Use this information to determine if the ant is going uphill or downhill (or neither) at $t = 5$ seconds.

Answer: The ant is going uphill.

5. The multivariable chain rule is a key tool in modern machine learning. In the big picture, neural networks "learn" parameters through an algorithm called *gradient descent*. This algorithm involves computing total derivatives of long chains of functions in high dimensions, which is in general extremely hard to do. The chain rule tells us that instead of contending with such a long chain of functions all at once, one can instead study each "layer" by itself, then combine everything with matrix multiplication, which is relatively easier. This is called *backpropagation*.

Answer: Right now this is just an interesting piece of information.

Extensions

6. Suppose we have a differentiable function $w = g(x, y)$ and x and y are differentiable functions of t and we know the following information.

$$g(1,0) = 1, \ g_x(1,0) = -2, \ g_y(1,0) = 2, \ g(-1,2) = 3, \ g_x(-1,2) = 1, \ g_y(-1,2) = -2,$$

$$x(2) = 1, \ y(2) = 0, \ x(1) = 1, \ y(1) = 3, \ x'(2) = 4, \ y'(2) = -1, \ x'(1) = 0, \ y'(1) = 2$$

If possible, find $\frac{dw}{dt}(1)$ and $\frac{dw}{dt}(2)$ or explain why the given information is not enough to do so. Which of these pieces of information would you not use at all to compute either value?

Answer: $\frac{dw}{dt}(2) = -10$. $\frac{dw}{dt}(1)$ cannot be computed from the given information because we do not know the values of g_x or g_y at $(x(1), y(1)) = (1, 3)$. We do not use the values of $g(1,0), g(-1,2), g_x(-1,2), g_y(-1,2)$.

7. Give an example of a nonconstant, differentiable function $f : \mathbb{R}^2 \to \mathbb{R}^2$ for which Df is **not** invertible at $(0,0)$, but invertible at $(1,0)$.

Answer: Many possible answers. One is

$$f(x,y) = \begin{bmatrix} x^2 \\ y \end{bmatrix}$$