1   The Chain Rule

If $z = g(y)$ and $y = f(x)$ and then the chain rule is a formula for the derivative of z with respect to x. If we let $z = h(x) = g(f(x))$, the chain rule formula is

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \text{ or } h'(x) = g'(y)f'(x) \text{ where } y = f(x).$$

The second formula shows where the derivatives are to be evaluated. In other words $\frac{dz}{dx}$ is evaluated at x, $\frac{dz}{dy}$ is evaluated at $y = f(x)$ and $\frac{dy}{dx}$ is evaluated at x.

For example, if $y = e^x$ and $z = \log y = \ln y$, then $z = x$ as $\log(e^x) = x$. This gives us

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = 1 \text{ or } \frac{dx}{dx} = \frac{dx}{dy}\frac{dy}{dx} = 1.$$

This formula needs to read carefully as it says that $\frac{dx}{dy}$, when $y = e^x$, is equal to the reciprocal of $\frac{dy}{dx}$ at x.

The typical chain rule in calculus of several variables is more complicated. For example, if $T = T(x, y)$ is viewed as the temperature at (x, y) and $\vec{r}(t) = (x(t), y(t))$ is a path or curve in the plane, then the temperature $T(x(t), (y(t))$ depends on t as the curve is traversed. This temperature at "time" t, is the temperature $T(x(t), y(t))$ at $(x(t), y(t))$, namely the temperature at the location of "moving point" $(x(t), y(t))$ at "time" t. Then the rate of change of $T = (x(t), y(t))$ is given by the formula

$$\frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial x}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right) = \nabla T \cdot \frac{d\vec{r}}{dt} = (\nabla T)(\vec{r}(t)) \cdot \frac{d\vec{r}}{dt}.$$

Here, the partial derivatives with respect to x and y are to be evaluated at $(x, y) = (x(t), y(t))$.

One might also have x and y themselves functions of several variables, say, $u, v,$ and w. This situation is given by a function $(x, y) = F(u, v, w) = (F_1(u, v, w), F_2(u, v, w))$. Then, if one fixes in turn all the variables but one, one gets three chain rule formulas

$$\frac{\partial T}{\partial u} = \frac{\partial T}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial u} = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}\right) \cdot \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\right)$$

$$\frac{\partial T}{\partial v} = \frac{\partial T}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial v} = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}\right) \cdot \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}\right)$$

$$\frac{\partial T}{\partial w} = \frac{\partial T}{\partial x}\frac{\partial x}{\partial w} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial w} = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}\right) \cdot \left(\frac{\partial x}{\partial w}, \frac{\partial y}{\partial w}\right).$$

These three chain rule formulas can all be "packaged" as one equation if one uses matrix multiplication. It involves the gradient of T and what is called the derivative (matrix) DF or F' of F. This is the matrix

$$DF = F' = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{bmatrix}.$$

Notice the pattern of the matrix, it follows the pattern of the variables (x, y) and (u, v, w). The three formulas for the three partial derivatives of T can be written as one matrix equation

$$\begin{bmatrix} \frac{\partial T}{\partial u} & \frac{\partial T}{\partial v} & \frac{\partial T}{\partial w} \end{bmatrix} = \begin{bmatrix} \frac{\partial T}{\partial x} & \frac{\partial T}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{bmatrix}.$$

Setting $p = (u, v, w)$, this can be written in the form

$$D(T \circ F)(p) = DT(F(p))DF(p) \text{ or } (T \circ F)'(p) = T'(F(p))F'(p)$$

where the sign ∘ stands for composition of functions. In this way the chain rule for functions of several variables looks exactly the same as the one variable case.

The derivative matrix DF is often called the **Jacobian matrix**. For example, when

$$(y_1, y_2, \cdots, y_m) = F(x_1, x_2, \cdots, x_n)$$

the Jacobian matrix (see Adams, p. 189 of the third edition) is the matrix

$$DF = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

In other words, the (i, j)-entry is $\frac{\partial y_i}{\partial x_j}$. When $m = n$ the determinant of the Jacobian matrix DF is called the **Jacobian** of F or the **Jacobian determinant** of F and is sometimes denoted by

$$\frac{\partial(y_1, y_2, \ldots, y_m)}{\partial(x_1, x_2, \ldots, x_n)}.$$

One important consequence of the matrix version of the chain rule is that if

$$(y_1, y_2, \cdots, y_n) = F(x_1, x_2, \cdots, x_n) \text{ and } (x_1, x_2, \cdots, x_n) = G(y_1, y_2, \cdots, y_n)$$

are two transformations such that $(x_1, x_2, \cdots, x_n) = G(F(x_1, x_2, \cdots, x_n))$ then the Jacobian matrices DF and DG are inverse to one another. This is because, if $I(x_1, x_2, \cdots, x_n) = (x_1, x_2, \cdots, x_n)$ then DI is the identity matrix $n \times n$ matrix I_n. Hence,

$$I_n = D(I) = D(F \circ G) = D(F)D(G).$$

Taylor's formula for a function of several variables is an immediate consequence of Taylor's formula in one variable and the chain rule. For example if $y = f(x)$ with $x = (x_1, x_2)$ and one wants the Taylor polynomial of order 2 of the function corresponding to the point $a = (a_1, a_2)$ let

$$g(t) = f(a + t(x - a)).$$

Then it follows that

$$g'(t) = \nabla f(a + t(x - a)) \cdot (x - a)$$

$$= \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(a + t(x - a))(x_i - a_i)$$

$$g''(t) = \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(a + t(x - a))(x_i - a_i)(x_j - a_j).$$

The Taylor polynomial of order 2 for g corresponding to $t = 0$ is

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2}g''(0).$$

If we evaluate this polynomial at $t = 1$ and observe that $g(1) = f(x)$, then

$$f(x) = f(a) + \sum_{i=1}^{2} \frac{\partial f}{\partial x_1}(a)(x_i - a_i) + \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i - a_i)(x_j - a_j).$$