The Chain Rule

For derivatives of functions of more than one variable, matrices play the role that numbers play for derivatives of functions of one variable. We'll see how that works for the Chain Rule.

First we'll recall the setup for functions of one variable.
Then we'll see the generalization.

Functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$.

The composition $h = g \circ f$ of the two functions is defined as follows. If a is an input to f, then $(g \circ f)(a) = (g(f(a))$. The Chain Rule holds that the derivative of the composition is the product of the derivatives:

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a) \ .$$

The chain rule for functions involving more variables is the same – the only difference is that now the derivative is a matrix instead of a number, and the product on the right side is a product of matrices.

Example: functions of one variable. Suppose $f(x) = 3x^2$ and $g(y) = y^4$. The $(g \circ f)(x) = (3x^2)^4 = 81x^8$, and $(g \circ f)'(x) = 8(81)x^7$. Then $f'(x) = 6x$ and $g'(y) = 4y^3$. At an input $x = 2$, we have

$$\begin{aligned}
\big(g'(f(2)\big)\big(f'(2)\big) &= \big(g'(12)\big)\big(f'(2)\big) = \big(4(12)^3\big)\big(6(2)\big) \\
&= 2^2(2^2 \cdot 3)^3(3 \cdot 2 \cdot 2) = 2^2(2^6 \cdot 3^3)(3 \cdot 2 \cdot 2) = 2^{10} \cdot 3^4 \\
\big(g \circ f\big)'(2) &= 8(81)2^7 = 2^3 \cdot (3^4)2^7 = 2^{10} \cdot 3^4 \ .
\end{aligned}$$

The two computations agree, as the Chain Rule tells us. (Note, by comparing factorizations I avoided the multipliction to show that $2^{10}3^4 = 82,944$. I hope.)

Example: functions $f:\mathbb{R}^2\to\mathbb{R}^2$, $g:\mathbb{R}^2\to\mathbb{R}$.
Let

$$\begin{aligned} f(x,y) &= (u,v) = (x^2y, xy^3) \\ g(u,v) &= z = u^2 - 3v \ . \end{aligned}$$

Then

$$\begin{aligned} g'(u,v) &= \begin{pmatrix} \partial z/\partial u & \partial z/\partial v \end{pmatrix} = \begin{pmatrix} 2u & -3 \end{pmatrix} \\ f'(x,y) &= \begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{pmatrix} = \begin{pmatrix} 2xy & x^2 \\ y^3 & 3xy^2 \end{pmatrix} \end{aligned}$$

Now consider the particular input $(x,y)=(1,2)$ to f. We have $f(1,2) = (1^2\cdot 2, 1\cdot 2^3) = (2,8)$. Let us compute $(g\circ f)'(1,2)$ using the Chain Rule:

$$\begin{aligned} (g'(u,v))(f'(x,y)) &= \begin{pmatrix} 2u & -3 \end{pmatrix} \begin{pmatrix} 2xy & x^2 \\ y^3 & 3xy^2 \end{pmatrix} \\ (g\circ f)'(1,2) = (g'(2,8))(f'(1,2)) &= \begin{pmatrix} 2(2) & -3 \end{pmatrix} \begin{pmatrix} 2(1)(2) & 1^2 \\ 2^3 & 3(1)2^2 \end{pmatrix} \\ &= \begin{pmatrix} 4 & -3 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 8 & 12 \end{pmatrix} = \begin{pmatrix} -8 & -32 \end{pmatrix} \end{aligned}$$

Let's see the alternate computation agrees. We have

$$\begin{aligned} (g\circ f)(x,y) &= z = u^2 - 3v = (x^2y)^2 - 3(xy^3) \\ &= x^4y^2 - 3xy^3 \\ (g\circ f)'(x,y) &= \begin{pmatrix} \partial z/\partial x & \partial z/\partial y \end{pmatrix} = \begin{pmatrix} 4x^3y^2 - 3y^3 & 2x^4y^2 - 9xy^2 \end{pmatrix} \\ (g\circ f)'(1,2) &= \begin{pmatrix} 16-24 & 8-36 \end{pmatrix} = \begin{pmatrix} -8 & -32 \end{pmatrix} \ . \end{aligned}$$

All is well.

Sample problem. Use the chain rule to compute $(g\circ f)'(1,2)$ if $f:\mathbb{R}^2\to\mathbb{R}^2$ and $g:\mathbb{R}^2\to\mathbb{R}^2$ are defined by $f(x,y)=(u,v)=(x^2y, xy^2)$ and $g(u,v) = (y,z) = (u^3, v^2u+u)$.