Have previously covered methods for differentiating single and multivariate functions.
Also covered calculating rapid derivatives and for navigating the gradients of high dimensional spaces.

The chain rule is another useful tool.
It fragments derivative problems into smaller problems called chains — the sum of the solutions to those chains is the product of the chain rule.

Imagine, the total derivative,

$f(x, y, z) = \sin x \, e^{yz^2}$, where the original variables x, y, z are the function of some other variable $\underline{t}$.

The derivative of $f(x, y, z)$ can be calculated

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

That is the sum of the chains relating $\underline{f}$-to-$\underline{t}$, through each of its variables.

Computers benefit from this approach — rather than substituting the t-values into the original equation and trying to derive the likely longer equation.

This total derivative process can be generalized using the following notation.

$f(x_1, x_2, \ldots, x_n) = f(x)$, where x is shorthand for params $\{x_i \mid 0 \le i \le n\}$.

Should think of x as a sequence of variables – a n-dimensional vector.

Each of the components of the x-vector are functions of t.

$$x_1(t) = \ldots$$
$$x_2(t) = \ldots$$
$$\ldots$$
$$x_n(t) = \ldots$$

What we want,

$$\frac{df}{dt}.$$

What we have,

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} \qquad \frac{dx}{dt} = \begin{bmatrix} dx_1/dt \\ dx_2/dt \\ \vdots \\ dx_n/dt \end{bmatrix}$$

The calculation,

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt}$$, the dot product.

$$= \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} \cdot \begin{bmatrix} dx_1/dt \\ dx_2/dt \\ \vdots \\ dx_n/dt \end{bmatrix}$$

$f(x(t))$ is another way of writing $f(x)$.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}$$ is the transposed Jacobian!

That is,

$$\frac{\partial f}{\partial x} = (J_f)^T$$

This Jacobian notation allows for the derivative of f relative to t to be written,

Where the dot product is equivalent to using the non-transposed Jacobian $\longrightarrow$

$$\frac{df}{dt} = J_f \frac{dx}{dt}.$$

The multivariate chain rule works for chains longer than one.

$f(x) = 5x$
$g(u) = 1-u$
$u(t) = t^2$

This can also be achieved with the chain rule.

$$\frac{df}{dt} = \frac{df}{dg}\frac{dg}{du}\frac{du}{dt} \Leftrightarrow 5(-1)(2t)$$
$$\Leftrightarrow -10t$$

$$f(t) = 5(1-t^2) \Leftrightarrow \frac{df}{dt} = -10t$$
$$= 5-5t^2$$

This technique also works for multivariate chains.

$f(x(u(t)))$: some multivariate function

$f(x) = f(x_1, x_2)$, takes a vector x as an input.

$x(u) = \begin{bmatrix} x_1(u_1, u_2) \\ x_2(u_1, u_2) \end{bmatrix}$, a vector-valued function which takes u (a vector) as its input.

$u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$, is a vector valued function that takes t (a scalar) as it's input.

The total derivative can again be written,

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u}\frac{du}{dt}$$

$$= \underbrace{\begin{bmatrix} \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \end{bmatrix}}_{J_f} \underbrace{\begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix}}_{J_x} \underbrace{\begin{bmatrix} \frac{du_1}{t} \\ \frac{du_2}{t} \end{bmatrix}}$$

to find $\frac{\partial x}{\partial u}$ must find the derivative of the two output variables with respect to the two input variables.

derivative vector u

(*) The matrix multiplication works such that a scalar is calculated.
$(1\times2)(2\times2)(2\times1) = 1$.