Scalar derivatives

Some notation for scalar derivatives:

- For a function $f \in \mathbb{R} \to \mathbb{R}$, we write $f' \in \mathbb{R} \to R$ for its derivative with respect to its argument. If the argument is called x, we can also write $\frac{d}{dx} f$. If the argument represents time, we sometimes write $\dot{f}$.
- If a function depends on more than one variable, we write $\frac{\partial}{\partial x} f$ or $\frac{\partial}{\partial y} f$ to indicate a *partial* derivative: the derivative with respect to one variable while holding the others constant.
- Second and higher derivatives are f'', $\ddot{f}$, $\frac{d^2}{dx^2} f$, or $\frac{\partial^2}{\partial x \partial y} f$.
- For a function f, we write $f\big|_{\hat{x}}$ or $f(x)\big|_{x=\hat{x}}$ to represent evaluation at $\hat{x}$. This means the same thing as $f(\hat{x})$ but is sometimes clearer: it lets us keep one name (x) for the variable we are differentiating, and another name ($\hat{x}$) for the value we are substituting at the end.

Scalar identities

Some of the most common identities for working with scalar derivatives:

- Differentiation and partial differentiation are linear operators: for example, $(af + bg)' = af' + bg'$.
- Chain rule: if we want $\frac{d}{dx} f(g(x))$, then we use

  $$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}$$

  (As a mnemonic, we can "cancel the dg" — but since $\frac{df}{dg}$ isn't really division, this is just a mnemonic.) Another way to write the same thing:

  $$\frac{d}{dx} f(g(x)) = f'(g(x))\, g'(x)$$

- Product rule:

  $$(fg)' = f'g + fg'$$

Common functions

Here are some useful derivatives of scalar functions. In each expression, x is the variable of interest; all other symbols represent constants.

- The derivative of a constant is zero: $\frac{d}{dx}a = 0$.
- The derivative of a monomial x^k is kx^{k-1}. This works even for negative and fractional values of k. One special case is x^0, where by convention we treat $0x^{-1}$ as equal to zero everywhere.
- The derivative of $\sin x$ is $\cos x$; the derivative of $\cos x$ is $-\sin x$.
- The derivative of e^{ax} is ae^{ax}. If we're using some other base b, we rewrite $b^x = e^{x \ln b}$ and then use the identity above.
- The derivative of $\ln x$ is x^{-1}. Again we can easily switch to another base: $\log_b x = \ln x / \ln b$.

Vector derivatives

It's also useful to think about functions that return vectors or take vectors as arguments. If f is a vector-valued function of a real argument, $f \in \mathbb{R} \to \mathbb{R}^n$, we can write it as a vector whose components are real-valued functions,

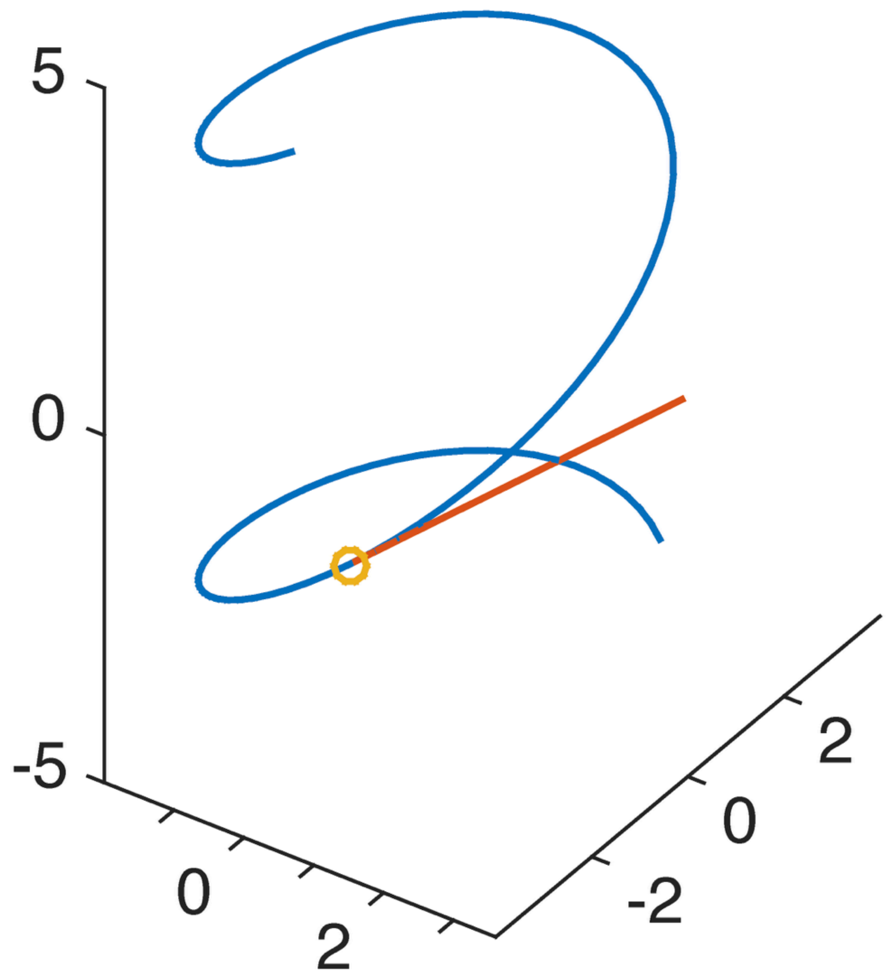
$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

Its derivative is then also a vector-valued function, of the same shape as f. Its components are the derivatives of the component functions:

$$\frac{d}{dx}f = \begin{pmatrix} \frac{df_1}{dx} \\ \frac{df_2}{dx} \\ \vdots \\ \frac{df_n}{dx} \end{pmatrix}$$

We can think of f as representing a curve in $\mathbb{R}^n$. The derivative $\frac{df}{dx}$ represents a tangent vector to this curve: the instantaneous velocity of a point moving along the curve as the argument x changes at a unit rate. The length of the tangent vector tells us the speed of the point, and the components tell us its direction.

Here's an example of a function in $\mathbb{R} \to \mathbb{R}^3$ and its derivative at a particular point:



Note that this plot doesn't show the argument x explicitly: instead it is implicit in the position of the point along the curve. If we wanted to show x explicitly, we could color the curve or add grid marks to show what values of x correspond to what values of $f(x)$.

More vector derivatives

If the function f has multiple inputs instead of multiple outputs, $f \in \mathbb{R}^n \to \mathbb{R}$, we can collect all of the arguments into a column vector:

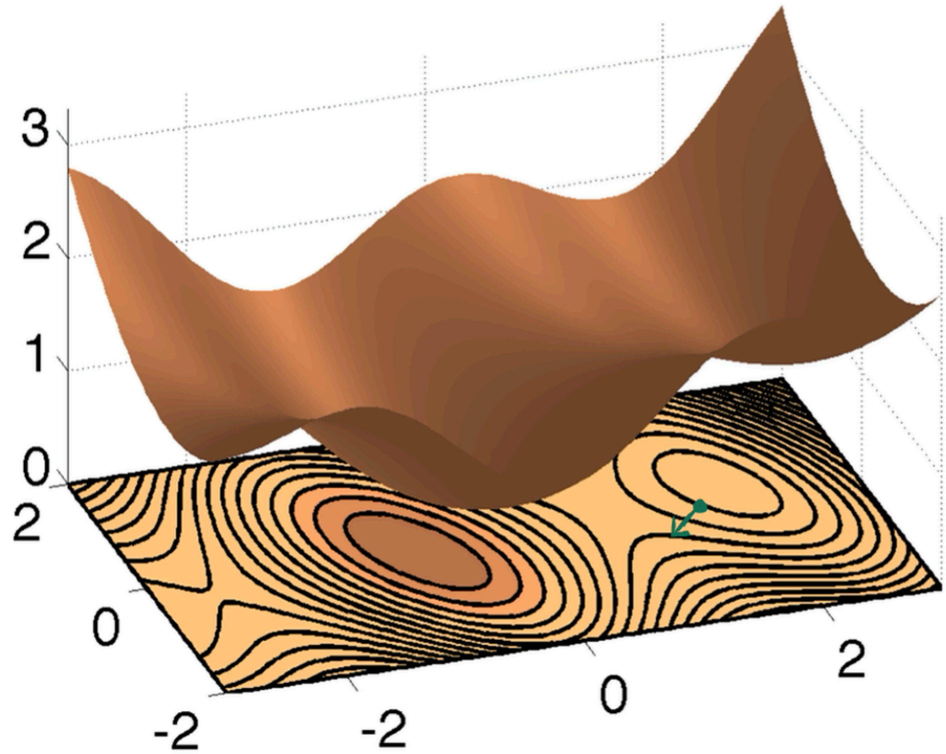
$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Then $\frac{df}{dx}$ means the row vector of partial derivatives of f:

$$\frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$

We can think of f as representing a surface in $\mathbb{R}^{n+1}$: the argument x varies across $\mathbb{R}^n$ while $f(x)$ determines the height. In this case the tangent vector tells us the direction of steepest increase of the function.

Here's an example of a function in $\mathbb{R}^2 \to \mathbb{R}$ together with its derivative at a point:



The derivative is the vector in $\mathbb{R}^2$ (shown in green at the bottom of the plot) that points in the direction of steepest increase. Note that it is orthogonal to a contour line.

Chain rule for vectors

With the above notation, the chain rule for vector functions looks just like it did for scalar functions. Suppose $f \in \mathbb{R}^n \to \mathbb{R}$ takes multiple arguments and $g \in \mathbb{R} \to \mathbb{R}^n$ returns multiple values, so that $f(g(x))$ makes sense. Then we have

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}$$

This looks just like the scalar chain rule (we "cancel the dg"). But now $\frac{df}{dg}$ is a row vector in $\mathbb{R}^{1\times n}$ and $\frac{dg}{dx}$ is a column vector in $\mathbb{R}^{n\times 1}$, so that when we multiply them we get their dot product. For clarity we can indicate the values of the arguments to each function:

$$\left.\frac{df}{dx}\right|_x = \left.\frac{df}{dg}\right|_{g(x)} \left.\frac{dg}{dx}\right|_x$$

If we write out the dot product, we get

$$\frac{df}{dx} = \sum_{i=1}^{n} \frac{\partial f}{\partial g_i} \frac{dg_i}{dx}$$

which may be familiar as the rule for calculating the *total derivative* of f with respect to x. In words, to calculate the change in f, we sum up the effects of all of the changes in all of the inputs to f.