9. The derivative

The derivative of a function represents the best linear approximation of that function. In one variable, we are looking for the equation of a straight line. We know a point on the line so that we only need to determine the slope.

Definition 9.1. *Let $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ be a function and let $a \in \mathbb{R}$ be a real number. f is **differentiable at** a, with derivative $\lambda \in \mathbb{R}$, if*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lambda.$$

To understand the definition of the derivative of a multi-variable function, it is slightly better to recast (9.1):

Definition 9.2. *Let $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ be a function and let $a \in \mathbb{R}$ be a real number. f is **differentiable at** a, with derivative $\lambda \in \mathbb{R}$, if*

$$\lim_{x \to a} \frac{f(x) - f(a) - \lambda(x - a)}{x - a} = 0.$$

We are now ready to give the definition of the derivative of a function of more than one variable:z

Definition 9.3. *Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a function and let $P \in \mathbb{R}^n$ be a point. f is **differentiable at** P, with derivative the $m \times n$ matrix A, if*

$$\lim_{Q \to P} \frac{f(Q) - f(P) - A\overrightarrow{PQ}}{\|\overrightarrow{PQ}\|} = 0.$$

We will write $Df(P) = A$.

So how do we compute the derivative? We want to find the matrix A. Suppose that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then

$$A\hat{e}_1 = A \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}$$

and

$$A\hat{e}_2 = A \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}.$$

In general, given an $m \times n$ matrix A, we get the jth column of A, simply by multiplying A by the column vector determined by $\hat{e}_j$.

So we want to know what happens if we approach P along the line determined by $\hat{e}_j$. So we take $\overrightarrow{PQ} = h\hat{e}_j$, where h goes to zero. In other words, we take $Q = P + h\hat{e}_j$. Let's assume that $h > 0$. So we consider the fraction

$$\begin{aligned}\frac{f(Q) - f(P) - A(h\hat{e}_j)}{\|\overrightarrow{PQ}\|} &= \frac{f(Q) - f(P) - A(h\hat{e}_j)}{h} \\ &= \frac{f(Q) - f(P) - hA\hat{e}_j}{h} \\ &= \frac{f(Q) - f(P)}{h} - A\hat{e}_j.\end{aligned}$$

Taking the limit we get the jth column of A,

$$A\hat{e}_j = \lim_{h \to 0} \frac{f(P + h\hat{e}_j) - f(P)}{h}.$$

Now $f(P + h\hat{e}_j) - f(P)$ is a column vector, whose entry in the ith row is

$$f_i(P+\hat{e}_j) - f_i(P) = f_i(a_1, a_2, \dots, a_{j-1}, a_j+h, a_{j+1}, \dots, a_n) - f_i(a_1, a_2, \dots, a_{j-1}, a_j, a_{j+1}, \dots, a_n).$$

and so for the expression on the right, in the ith row, we have

$$\lim_{h \to 0} \frac{f_i(P + h\hat{e}_j) - f_i(P)}{h}.$$

Definition 9.4. *Let $g \colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function and let $P \in \mathbb{R}^n$. The **partial derivative** of f at $P = (a_1, a_2, \dots, a_n)$, with respect to x_j is the limit*

$$\left.\frac{\partial f}{\partial x_j}\right|_P = \lim_{h \to 0} \frac{g(a_1, a_2, \dots, a_j + h, \dots, a_n) - g(a_1, a_2, \dots, a_n)}{h}.$$

Putting all of this together, we get

Proposition 9.5. *Let $f \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a function.*

If f is differentiable at P, then $Df(P)$ is the matrix whose (i, j) entry is the partial derivative

$$\left.\frac{\partial f_i}{\partial x_j}\right|_P.$$

Example 9.6. *Let $f \colon A \longrightarrow \mathbb{R}^2$ be the function*

$$f(x, y, z) = (x^3y + x\sin(xz), \log xyz).$$

Here $A \subset \mathbb{R}^3$ is the first octant, the locus where x, y and z are all positive. Supposing that f is differentiable at P, then the derivative is given by the matrix of partial derivatives,

$$Df(P) = \begin{pmatrix} 3x^2y + \sin(xz) + xz\cos(xz) & x^3 & x^2\cos(xz) \\ \frac{1}{x} & \frac{1}{y} & \frac{1}{z} \end{pmatrix}.$$

Definition 9.7. *Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable function. Then the derivative of f at P, $Df(P)$ is a row vector, which is called the **gradient** of f, and is denoted $(\nabla f)\big|_P$,*

$$(\frac{\partial f}{\partial x_1}\bigg|_P, \frac{\partial f}{\partial x_2}\bigg|_P, \dots, \frac{\partial f}{\partial x_n}\bigg|_P).$$

The point $(x_1, x_2, \dots, x_n, x_{n+1})$ lies on the graph of $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ if and only if $x_{n+1} = f(x_1, x_2, \dots, x_n)$.

The point $(x_1, x_2, \dots, x_n, x_{n+1})$ lies on the **tangent hyperplane** of $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ at $P = (a_1, a_2, \dots, a_n)$ if and only if

$$x_{n+1} = f(a_1, a_2, \dots, a_n) + (\nabla f)\big|_P \cdot (x_1 - a_1, x_2 - a_2, \dots, x_n - a_n).$$

In other words, the vector

$$(\frac{\partial f}{\partial x_1}\bigg|_P, \frac{\partial f}{\partial x_2}\bigg|_P, \dots, \frac{\partial f}{\partial x_n}\bigg|_P, -1),$$

is a normal vector to the tangent hyperplane and of course the point $(a_1, a_2, \dots, a_n, f(a_1, a_2, \dots, a_n))$ is on the tangent hyperplane.

Example 9.8. *Let*

$$D = \{\, (x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 < r^2 \,\},$$

the open ball of radius r, centred at the origin.

Let $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ be the function given by

$$f(x, y) = \sqrt{r^2 - x^2 - y^2}.$$

Then

$$\frac{\partial f}{\partial x} = \frac{-2x/2}{\sqrt{r^2 - x^2 - y^2}} = \frac{-x}{\sqrt{r^2 - x^2 - y^2}},$$

and so by symmetry,

$$\frac{\partial f}{\partial y} = \frac{-y}{\sqrt{r^2 - x^2 - y^2}} = \frac{-x}{\sqrt{r^2 - x^2 - y^2}},$$

At the point (a, b), the gradient is

$$(\nabla f)\big|_{(a,b)} = \frac{-1}{\sqrt{r^2 - a^2 - b^2}}(a, b).$$

So the equation for the tangent plane is

$$z = f(a, b) - \frac{1}{\sqrt{r^2 - a^2 - b^2}}(a(x - a) + b(x - b)).$$

For example, if $(a, b) = (0, 0)$, then the tangent plane is

$$z = r,$$

as expected.