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*

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

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,

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

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.

10. More about derivatives

The main result is:

Theorem 10.1. *Let $A \subset \mathbb{R}^n$ be an open subset and let $f\colon A \longrightarrow \mathbb{R}^m$ be a function.*

If the partial derivatives

$$\frac{\partial f_i}{\partial x_j},$$

exist and are continuous, then f is differentiable.

We will need:

Theorem 10.2 (Mean value theorem)**.** *Let $f\colon [a,b] \longrightarrow \mathbb{R}$ is continuous and differentiable at every point of (a,b), then we may find $c \in (a,b)$ such that*

$$f(b) - f(a) = f'(c)(b-a).$$

Geometrically, (10.2) is clear. However it is surprisingly hard to give a complete proof.

Proof of (10.1). We may assume that $m = 1$. We only prove this in the case when $n = 2$ (the general case is similar, only notationally more involved). So we have

$$f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}.$$

Suppose that $P = (a,b)$ and let $\overrightarrow{PQ} = h_1\hat{\imath} + h_2\hat{\jmath}$. Let

$$P_0 = (a,b) \qquad P_1 = (a+h_1, b) \qquad \text{and} \qquad P_2 = (a+h_1, b+h_2) = Q.$$

Now

$$f(Q) - f(P) = [f(P_2) - f(P_1)] + [f(P_1) - f(P_0)].$$

We apply the Mean value theorem twice. We may find Q_1 and Q_2 such that

$$f(P_1) - f(P_0) = \frac{\partial f}{\partial x}(Q_1)h_1 \qquad \text{and} \qquad f(P_2) - f(P_1) = \frac{\partial f}{\partial y}(Q_2)h_2.$$

Here Q_1 lies somewhere on the line segment P_0P_1 and Q_2 lies on the line segment P_1P_2. Putting this together, we get

$$f(Q) - f(P) = \frac{\partial f}{\partial x}(Q_1)h_1 + \frac{\partial f}{\partial y}(Q_2)h_2.$$

Thus

$$
\begin{aligned}
\frac{|f(Q)-f(P)-A\cdot\overrightarrow{PQ}|}{\|\overrightarrow{PQ}\|} &= \frac{|(\frac{\partial f}{\partial x}(Q_1)-\frac{\partial f}{\partial x}(P))h_1+(\frac{\partial f}{\partial y}(Q_2)-\frac{\partial f}{\partial y}(P))h_2|}{\|\overrightarrow{PQ}\|} \\
&\leq \frac{|(\frac{\partial f}{\partial x}(Q_1)-\frac{\partial f}{\partial x}(P))h_1|}{\|\overrightarrow{PQ}\|}+\frac{|(\frac{\partial f}{\partial y}(Q_2)-\frac{\partial f}{\partial y}(P))h_2|}{\|\overrightarrow{PQ}\|} \\
&\leq \frac{|(\frac{\partial f}{\partial x}(Q_1)-\frac{\partial f}{\partial x}(P))h_1|}{|h_1|}+\frac{|(\frac{\partial f}{\partial y}(Q_2)-\frac{\partial f}{\partial y}(P))h_2|}{|h_2|} \\
&= |(\frac{\partial f}{\partial x}(Q_1)-\frac{\partial f}{\partial x}(P))|+|(\frac{\partial f}{\partial y}(Q_2)-\frac{\partial f}{\partial y}(P))|.
\end{aligned}
$$

Note that as Q approaches P, Q_1 and Q_2 both approach P as well. As the partials of f are continuous, we have

$$\lim_{Q\to P}\frac{|f(Q)-f(P)-A\cdot\overrightarrow{PQ}|}{\|\overrightarrow{PQ}\|}\leq \lim_{Q\to P}(|(\frac{\partial f}{\partial x}(Q_1)-\frac{\partial f}{\partial x}(P))|+|(\frac{\partial f}{\partial y}(Q_2)-\frac{\partial f}{\partial y}(P))|)=0.$$

Therefore f is differentiable at P, with derivative A. □

Example 10.3. *Let $f\colon A\longrightarrow\mathbb{R}$ be given by*

$$f(x,y)=\frac{x}{\sqrt{x^2+y^2}},$$

where $A=\mathbb{R}^2-\{(0,0)\}$. Then

$$\frac{\partial f}{\partial x}=\frac{(x^2+y^2)^{1/2}-x(2x)(1/2)(x^2+y^2)^{-1/2}}{x^2+y^2}=\frac{y^2}{(x^2+y^2)^{3/2}}.$$

Similarly

$$\frac{\partial f}{\partial y}=-\frac{xy}{(x^2+y^2)^{3/2}}.$$

Now both partial derivatives exist and are continuous, and so f is differentiable, with derivative the gradient,

$$\nabla f=(\frac{y^2}{(x^2+y^2)^{3/2}},-\frac{xy}{(x^2+y^2)^{3/2}})=\frac{1}{(x^2+y^2)^{3/2}}(y^2,-xy).$$

Lemma 10.4. *Let $A=(a_{ij})$ be an $m\times n$ matrix.*

If $\vec{v}\in\mathbb{R}^n$ then

$$\|A\vec{v}\|\leq K\|\vec{v}\|,$$

where

$$K=(\sum_{i,j}a_{ij}^2)^{1/2}.$$

Proof. Let $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ be the rows of A. Then the entry in the ith row of $A\vec{v}$ is $\vec{a}_i \cdot \vec{v}$. So,

$$\begin{aligned}
\|A\vec{v}\|^2 &= (\vec{a}_1 \cdot \vec{v})^2 + (\vec{a}_2 \cdot \vec{v})^2 + \dots + (\vec{a}_n \cdot \vec{v})^2 \\
&\leq \|\vec{a}_1\|^2\|\vec{v}\|^2 + \|\vec{a}_2\|^2\|\vec{v}\|^2 + \dots + \|\vec{a}_n\|^2\|\vec{v}\|^2 \\
&= (\|\vec{a}_1\|^2 + \|\vec{a}_2\|^2 + \dots + \|\vec{a}_n\|^2)\|\vec{v}\|^2 \\
&= K^2\|\vec{v}\|^2.
\end{aligned}$$

Now take square roots of both sides. □

Theorem 10.5. *Let $f \colon A \longrightarrow \mathbb{R}^m$ be a function, where $A \subset \mathbb{R}^n$ is open.*

If f is differentiable at P, then f is continuous at P.

Proof. Suppose that $Df(P) = A$. Then

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

This is the same as to require

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

But if this happens, then surely

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

So

$$\begin{aligned}
\|f(Q) - f(P)\| &= \|f(Q) - f(P) - A \cdot \overrightarrow{PQ} + A \cdot \overrightarrow{PQ}\| \\
&\leq \|f(Q) - f(P) - A \cdot \overrightarrow{PQ}\| + \|A \cdot \overrightarrow{PQ}\| \\
&\leq \|f(Q) - f(P) - A \cdot \overrightarrow{PQ}\| + K\|\overrightarrow{PQ}\|.
\end{aligned}$$

Taking the limit as Q approaches P, both terms on the RHS go to zero, so that

$$\lim_{Q \to P} \|f(Q) - f(P)\| = 0,$$

and f is continuous at P. □