Derivatives of multivariable functions

Suppose there is a function $f : \mathbb{R}^2 \to \mathbb{R}$. What is the derivative of f at (a, b)?

We can consider this function in two-dimensional planes. Fix $y = b$, which gives us $f(\cdot, b) : \mathbb{R} \to \mathbb{R}$. We can find the derivative of this function $f'(\cdot, b)$, and the slope of the tangent line at (a, b), $f'(a, b)$. Equally, we can fix $x = a$, and do the same operation on $f(a, \cdot) : \mathbb{R} \to \mathbb{R}$, giving us the derivative $f'(a, \cdot)$ and once again the slope $f'(a, b)$. This gives us two lines that are orthogonal, since the plane where a is fixed and the plane where b is fixed are orthogonal, and we have lines given by slopes in these two planes. These two lines define a plane of their own, which is the slope of the function f at (a, b). The two orthogonal lines are called the **partial derivatives** of f.

The **total derivative** of a function $f : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map $T_x : \mathbb{R}^m \to \mathbb{R}^n$. A partial derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $\vec{a} = (a_1, a_2 \cdots a_n)$ is the derivative of the function $f(a_1, a_2 \cdots a_i \cdots a_n) : \mathbb{R} \to \mathbb{R}$, where the parameter a_i is fixed for some i. This is denoted f_i, f_{x_i} or $\frac{\partial f}{\partial x_i}$. Note that the partial derivative is a map $\mathbb{R}^n \to \mathbb{R}$.

For example, consider $f(x, y) = x^2 + y^3$. $f_x(a, b) = 2a$, and $f_y(a, b) = 3b^2$. Therefore, the total derivative of $f(x, y)$ is $(f_x, f_y) = (2x, 3y^2) : \mathbb{R}^2 \to \mathbb{R}$.

However, instead of knowing the derivative at (a, b) along the planes parallel to x and y, we want to know the derivative on some arbitrary plane. This is called a **directional derivative**, for some direction vector $\vec{h} = (h_1, h_2)$. Using the definition of the derivative, we have $f_{\vec{h}} = \lim_{t\to\infty} \frac{f((a,b)+t\vec{h})-f(a,b)}{t\|\vec{h}\|}$. For $f(x, y) = x^2 + y^3$, we have $f_{\vec{h}} = \lim_{t\to\infty} \frac{(a+th_1)^2+(b+th_2)^3-a^2-b^3}{t\|\vec{h}\|} = \lim_{t\to\infty} \frac{1}{\|\vec{h}\|}(2h_1 a + th_1^2 + 3h_2 b^2 + t * \cdots)$. The factor on t isn't important since t tends to 0. Therefore, we have the formula for the directional derivative with regards to $\vec{h}$: $\frac{1}{\|\vec{h}\|}(f_x, f_y) \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$.

Differentiability

A function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{x}$ iff all partial derivatives exist and are continuous at $\vec{x}$.

We can also say that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{a}$ if there is a "good" linear approximation $f(\vec{a} + \vec{h}) \approx f(\vec{a}) + l(\vec{h}$, where l is a linear function. Formally, this means there exists l such that $\lim_{\vec{h}\to\vec{0}} \frac{f(\vec{a}+\vec{h})-f(\vec{a})-l(\vec{h})}{\|\vec{h}\|} = 0$. If this function is differentiable, then $l = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{a}) & \frac{\partial f}{\partial x_2}(\vec{a}) & \cdots & \frac{\partial f}{\partial x_n}(\vec{a}) \end{bmatrix} = \nabla f(\vec{a})$.

To recap: if all $f_{x_i}(\vec{x})$ exist and are continuous in a neighbourhood of $\vec{a}$, then f is differentiable at $\vec{a} \in \mathbb{R}^n$, then f is differentiable at vva.

\begin{proof} A function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{a}$ if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}$ such that $\lim_{\|\vec{h}\|\to 0} \frac{\|f(\vec{a}+\vec{h})-f(\vec{a})-T(\vec{h})\|}{\|\vec{h}\|} = 0$.

Let $\vec{h} = (h_1, h_2 \cdots h_n)$, and let $\vec{v}^i = (h_1, h_2 \cdots h_i, 0 \cdots 0)$. $f(\vec{a} + \vec{h}) - f(\vec{a}) = \sum_{i=1}^{n} \left(f(\vec{a} + \vec{v}^i) - f(\vec{a} + \vec{v}^{i-1})\right) = \sum_{i=1}^{n} \left(f(\vec{a} + \vec{v}^{i-1} + h_i \vec{e}_i) - f(\vec{a} + \vec{v}^{i-1})\right)$. Let $g(t) = f(\vec{a} + \vec{v}^{i-1} + t \cdot \vec{e}_i) : \mathbb{R} \to \mathbb{R}$. $g'(t) = \lim_{\alpha\to 0} \frac{g(t+\alpha)-g(t)}{\alpha} = \lim_{\alpha\to 0} \frac{f(\vec{a}+\vec{v}^{i-1}+(t+\alpha)\vec{e}_i)-f(\vec{a}+\vec{v}^{i-1}+t\vec{e}_i}{t} = f_{x_i}(\vec{a} + \vec{v}^{i-1} + t\vec{e}_i)$. This eixists and is continuous for $\vec{v}^{i-1}$ and t small enough by assumption, in particular $t \in [0, h_i]$. By the mean value theorem, there exists t_i such that $\frac{g(h_i)-g(0)}{h_i} = g'(t_i)$. We can do this for every co-ordinate: $\forall\ i \exists t_i : f(\vec{a}+\vec{v}^{i-1}+h_i\vec{e}_i)-f(\vec{a}+\vec{v}^{i-1}) = h_i f_{x_i}(\vec{a} + \vec{v}^{i-1} + t_i\vec{e}_i)$. Therefore, the sum from before is equal to $\sum_{i=1}^{n} h_i f_{x_i}(\vec{a}+\vec{v}^{i-1}+t_i\vec{e}_i)$. Define $T(\vec{h}) = \sum f_{x_i}(\vec{a})h_i$. Now we can say that $\lim_{\|\vec{h}\|\to 0} \frac{\|f(\vec{a}+\vec{h})-f(\vec{a})-T(\vec{h})\|}{\|\vec{h}\|} = \lim_{\|\vec{h}\|\to 0} \frac{\|f(\vec{a}+\vec{h})-f(\vec{a})-\sum f_{x_i}(\vec{a})h_i\|}{\|\vec{h}\|} = \lim_{\|\vec{h}\|\to 0} \frac{\|\sum_{i=1}^{n}(h_i f_{x_i}(\vec{a}+\vec{v}^{i-1}+t_i\vec{e}_i)-f_{x_i}(\vec{a})h_i)\|}{\|\vec{h}\|}$. Fix $\epsilon > 0$; we can assume that $\left\|\vec{h}\right\|$ is small enough that $\forall\ i |f_{x_i}(\vec{a} + \vec{v}^{i-1} + t_i\vec{e}_i) - f_{x_i}(\vec{a})| < \frac{\epsilon}{n}$. Therefore, $\lim_{\|\vec{h}\|\to 0} \frac{\|\sum_{i=1}^{n}(h_i f_{x_i}(\vec{a}+\vec{v}^{i-1}+t_i\vec{e}_i)-f_{x_i}(\vec{a})h_i)\|}{\|\vec{h}\|} \le \sum_{i=1}^{n} \frac{1}{\|\vec{h}\|} |h_i f_{x_i}(\vec{a} + \vec{v}^{i-1} + t_i\vec{e}_i) - f_{x_i}(\vec{a})h_i|$. Now we have two cases. If $h_i = 0$, this term will be 0, and we are done. If $h_i \neq 0$, then the term in the sum is less than $\frac{1}{|h_i|}$, so the whole sum is less than or equal to $\sum_{i=1}^{n} \frac{1}{|h_i|} |h_i f_{x_i}(\vec{a} + \vec{v}^{i-1} + t_i\vec{e}_i) - f_{x_i}(\vec{a})h_i| = \sum_{i=1}^{n} |f_{x_i}(\vec{a} + \vec{v}^{i-1} + t_i\vec{e}_i) - f_{x_i}(\vec{a})| < n \cdot \frac{\epsilon}{n} = \epsilon$. /end{proof}

Tangent planes

In two dimensions, a function $f : \mathbb{R} \to \mathbb{R}$ had tangent lines to it, but a three-dimensional function $f : \mathbb{R}^2 \to \mathbb{R}$ has tangent planes. Take for example the function $f(x, y) = x^2 - y^3$. To find the tangent plane at $(1, 1)$ we first fix $y = 1$: $f(x, 1) = x^2 - 1$. Then, the corresponding partial derivative is $f_x(x, 1) = 2x$, and $f_x(1, 1) = 2$. Similarly, we fix $x = 1$: $f(1, y) = 1 - y^3$. Then the partial derivative is $f_y(1, y) = -3y^2$ and $f_y(1, 1) = -3$. Thus, a plane parallel to the tangent plane is the unique plane that contains the lines $z = 2x$ (when $y = 0$ is fixed) and $z = -3y$ (when $x = 0$ is fixed). This plane is $z = 2x - 3y$. It can be written as a dot product of two vectors: let $\vec{x} = \begin{bmatrix} x & y & z \end{bmatrix}$ and $\vec{u} = \begin{bmatrix} 2 & -3 & -1 \end{bmatrix}$, and the plane is $\vec{u} \cdot \vec{x}$. Then, the equation of the tangent plane is $\vec{u} \cdot \vec{x} = \begin{bmatrix} a & b & f(a, b) \end{bmatrix} \cdot \vec{x}$, for any a and b.

In general, the equation of the tangent plane of $f : \mathbb{R}^n \to \mathbb{R}$, $\begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} \end{bmatrix} \mapsto x_n$ at $\vec{a}\begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} \end{bmatrix}$ is $\begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & f(\vec{a}) \end{bmatrix} \cdot \begin{bmatrix} \nabla f(\vec{a}) & -1 \end{bmatrix} =$

$\begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \end{bmatrix} \cdot \begin{bmatrix} \nabla f(\vec{a}) & -1 \end{bmatrix}$. f is maximised or minimised whenever the tangent plane is flat, i.e. the orthogonal vector to the plane is $\begin{bmatrix} 0 & 0 & \cdots & -1 \end{bmatrix}$, which happens exactly when $\nabla f = \vec{0}$.