Derivatives of a Function of Several Variables

To move smoothly from derivatives of functions of one variable to derivatives of function of several variables, it is necessary to recast derivatives as linear maps. But what is a linear map? We start with the easiest one, namely a linear map from $\mathbb{R}$ into $\mathbb{R}$. It is a function whose graph is a straight line through the origin.

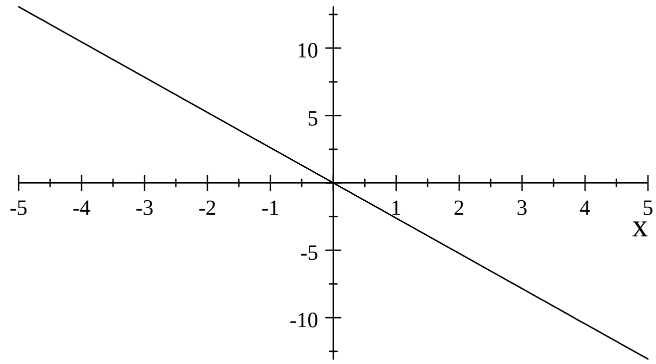


Figure (i). Graph of some linear function $L : \mathbb{R} \to \mathbb{R}$

Therefore it has a formula $L(x) = mx$ where m is a constant. Clearly, it satisfies the following two conditions: for any real numbers x and y,

$$L(x+y) = L(x) + L(y) \quad \text{and if } c \text{ is a constant then} \quad L(cx) = cL(x)$$

These are the very conditions used to define an arbitrary linear map. Thus, climbing one dimension higher, a function L from $\mathbb{R}^2$ into $\mathbb{R}$ is called a linear map if, for all pairs (h,k) and (x,y) in $\mathbb{R}^2$ and any constant c,

$$L((h,k)+(x,y)) = L(h,k) + L(x,y) \quad \text{and} \quad L(c(x,y)) = cL(x,y)$$

It turns out that the graph of such a function is a plane that pass through the origin $(0,0,0)$ and its formula has the form

$$L(x,y) = ax + by$$

where a and b are constants. This may be written as a product of the row matrix $\begin{pmatrix} a & b \end{pmatrix}$ and the column matrix $\begin{pmatrix} x \\ y \end{pmatrix}$, that is

$$L(x,y) = ax + by = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The 1×2 matrix $\begin{pmatrix} a & b \end{pmatrix}$ is called the matrix of L, (in the standard bases for $\mathbb{R}^2$ and $\mathbb{R}$).

Example 1 *Let $L(x,y) = 3x - 2y$. Then L is a linear map since*

$$\begin{aligned} L((h,k)+(x,y)) &= L(h+x, y+k) = 3(h+x) - 2y + k) = (3h-2k) + (3x-2y) \\ &= L(h,k) + L(x,y) \end{aligned}$$

and

$$L(c(x,y)) = L(cx, cy) = 3cx - 2cy = c(3x-2y) = cL(x,y)$$

It has a 1×2 matrix which is $\begin{pmatrix} 3 & -2 \end{pmatrix}$.

Going up one dimension higher in the range, a function L from $\mathbb{R}^2$ into $\mathbb{R}^2$ is linear if, for all pairs (h,k) and (x,y) in $\mathbb{R}^2$ and any constant c,

$$L((h,k)+(x,y)) = L(h,k)+L(x,y) \quad \text{and} \quad L(c(x,y)) = cL(x,y)$$

Example 2 *Let $L:\mathbb{R}^2 \to \mathbb{R}^2$ be defined by $L(x,y) = (4x+3y, -5x+y)$. Then L is linear since*

$$\begin{aligned} L((h,k)+(x,y)) &= L(h+x, k+y) = (4\,(h+x)+3(k+y), -5(h+x)+(k+y) \\ &= (4h+3k, -5h+k)+(4x+3y, -5x+y) = L(h,k)+L(x,y) \end{aligned}$$

and

$$L(c(x,y)) = L(cx,cy) = (4cx+3cy, -5cx+cy) = c(4x+3y, -5x+y) = cL(x,y)$$

In general, a linear map $L:\mathbb{R}^2 \to \mathbb{R}^2$ has the form

$$L(x,y) = (ax+by, cx+dy)$$

where a, b, c and d are constants. If we agree to write the image $(ax+by, cx+dy)$ as a column

$$\begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

instead of the row $\begin{pmatrix} ax+by & cx+dy \end{pmatrix}$ then $L(x,y)$ is the matrix product

$$L(x,y) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

As you would expect, the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is called the matrix for L, (in the standard bases for $\mathbb{R}^2$).

Example 3 *The linear map in Example 2 may be written in the matrix product form as*

$$L(x,y) = \begin{pmatrix} 4 & 3 \\ -5 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Its matrix is $\begin{pmatrix} 4 & 3 \\ -5 & 1 \end{pmatrix}$

Now consider a function $f(x)$ of one variable and a point c in its domain. Intuitively, f has a derivative at c if we can draw a tangent to its graph at $(c, f(c))$. The slope of the tangent is the number

$$f'(c) = \lim_{h \to 0} \frac{f(c+h)-f(c)}{h}$$

and its equation is

$$T(x) = (x-c)\,f'(c) + f(c).$$

As the figure below shows, when x is close to c, the point $(x, f(x))$ on the graph of f is close to the point $(x, T(x))$ on the tangent line.

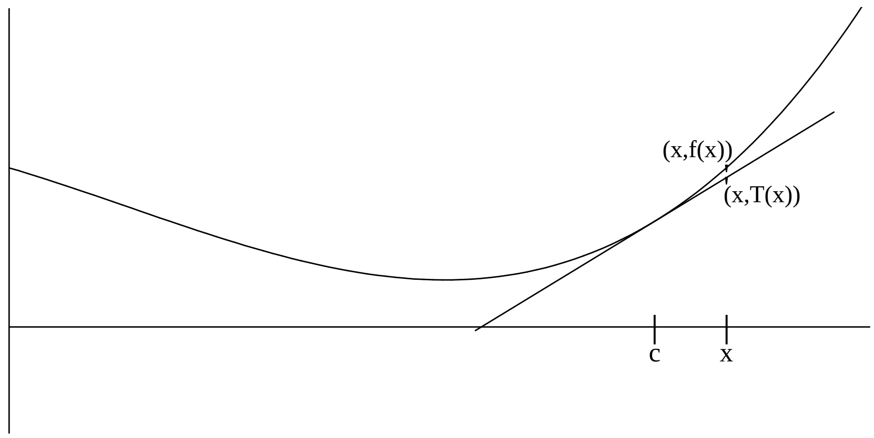


Figure (ii). $(x, f(x))$ is close to $(x, T(x))$

Therefore

$$f(x) \simeq T(x) = (x - c)\, f'(c) + f(c) \tag{1}$$

It is convenient to write x as $x = c + h$ and to subtract $f(c)$ from both sides of (1). The result is

$$f(c+h) - f(c) \simeq h f'(c)$$

This says that the difference $f(c+h) - f(c)$ can be approximated by the linear map $L(h) = f'(c)h$. The error in the approximation is the number

$$f(c+h) - f(c) - h f'(c)$$

To get an idea of how small it is, we appeal to a Taylor's theorem which asserts that there is a number θ between c and $c + h$ such that

$$f(c+h) = f(c) + h f'(c) + \tfrac{1}{2} f''(\theta) h^2$$

Therefore the error term is $f(c+h) - f(c) - h f'(c) = \frac{1}{2} f''(\theta) h^2$. When h is small, (with an absolute value less than 1), then h^2 is much smaller than h. Therefore the error term is very much smaller than h. In fact it is so much smaller than h that

$$\lim_{h \to 0} \frac{|\text{error term}|}{|h|} = 0.$$

These observations suggest that we may recast the differentiability of a function of one variable as follows:

Definition 4 *A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at a point $c \in A$ if $f(c+h) - f(c)$ can be approximated by a linear map $L(h)$, in such a way that when h is close to zero, the error term $|f(c+h) - f(c) - L(h)|$ is small compared to h. More precisely, f is differentiable at c if there is a linear map L such that*

$$\lim_{h \to 0} \frac{|f(c+h) - f(c) - L(h)|}{h} = 0$$

When convenient, we will call L in Definition 4 the linear approximator of $f(c+h) - f(c)$ at c.

Definition 5 *The linear approximator of $f(c+h) - f(c)$ is called the derivative of f at c and it is denoted by $Df(c)$.*

One may read $Df(c)$ as "the derivative of f at c". This notation has the advantage of pointing out that the linear approximator depends on the point c in the domain of f.

Example 6 *Let $f(x) = x^3 - 4x^2 + 1$ and $c = -1$. The tangent to its graph at $(-1, -4)$ has slope $f'(-1) = 11$. Therefore the linear approximator of $f(-1+h) - f(-1)$ at $x = -1$ is $Df(-1)$ with formula*

$$Df(-1)(h) = 11h.$$

We can verify directly that the error $|f(-1+h) - f(-1) - Df(-1)(h)|$ in the approximation is small compared to h. For

$$\begin{aligned} f(-1+h) - f(-1) &= \left[(-1+h)^3 - 4(-1+h)^2 + 1\right] - \left[(-1)^3 - 4(-1)^2 + 1\right] \\ &= h^3 - 4h^2 + 11h \end{aligned}$$

Therefore the error in approximation is $\left|h^3 - 4h^2 + 11h - 11h\right| = \left|h^3 - 4h^2\right|$ and

$$\lim_{h \to 0} \frac{\left|h^3 - 4h^2\right|}{|h|} = \lim_{h \to 0} \left|h^2 - 4h\right| = 0.$$

If we choose a different point, say $c = 4$, we are bound to get a different approximator, because the tangent at the new point need not be parallel to the tangent at $c = -1$. Indeed the linear approximator of $f(4+h) - f(4)$ at $x = 4$ is $Df(4)$ with formula $Df(4)(h) = 12h$, (since $f'(4) = 12$). In general, the linear approximator of $f(x+h) - f(x)$ at x is $Df(x)$ with formula

$$Df(x)(h) = (3x^2 - 8x)h$$

Derivative of a Real Valued Function of Two Variables

At the intuitive level, a real valued function g of one variable is differentiable at a point b if we can draw a tangent line to its graph at $(b, f(b))$. The graph of a function of one variable is a curve. When we climb one step higher to functions of two variables, we find that the graph of a real valued function $f(x, y)$ is a surface in space; which suggests that a *tangent line to a curve* should be upgraded to a *tangent plane to a surface*. Therefore f should be differentiable at a point (c, d) if we can draw a tangent plane to its graph at $(c, d, d(c, d))$.

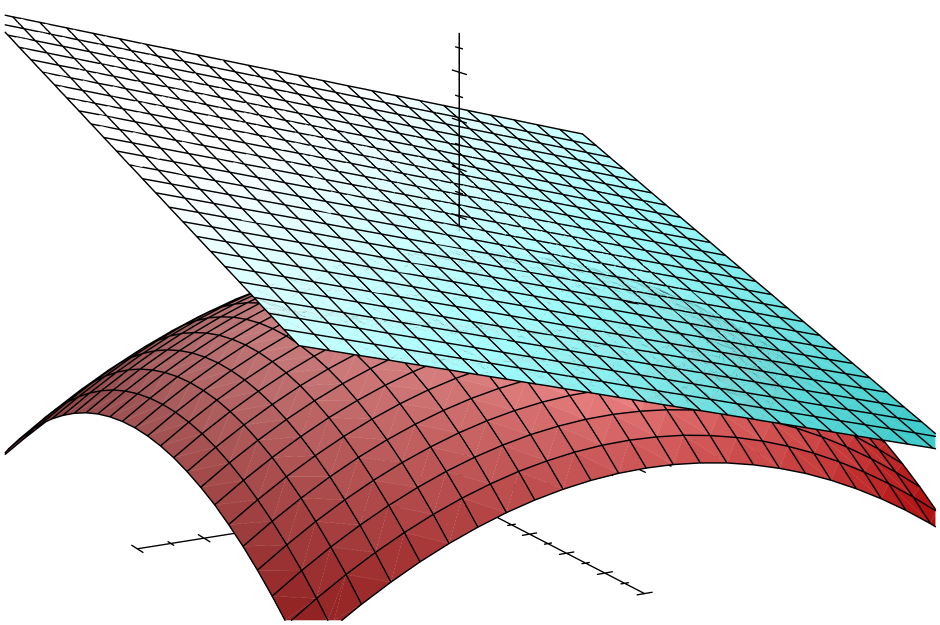


Figure (iii). A tangent plane at $P(c, d, f(c, d))$

To go beyond the intuitive definition of a differentiable function g of one variable, we determined the equation of the tangent at $(b, f(b))$ and used it to deduce that g is differentiable at b if the expression $f(b+h) - f(b)$

can be approximated by a linear mapping. We have to do the same for a function of two variables; i.e. we have to determine the equation of the tangent plane at $(c, d, f(c, d))$ and see what it suggests. To this end, consider the curves $\mathbf{c}_1(x) = (x, d, f(x, d))$ and $\mathbf{c}_2(y) = (c, y, f(c, y))$, (shown in the figure below), in the graph of f. They both pass through $(c, d, f(c, d))$. Furthermore, every tangent vector to $\mathbf{c}_1$ and every tangent vector to $\mathbf{c}_2$ at $(c, d, f(c, d))$ is in the tangent plane.

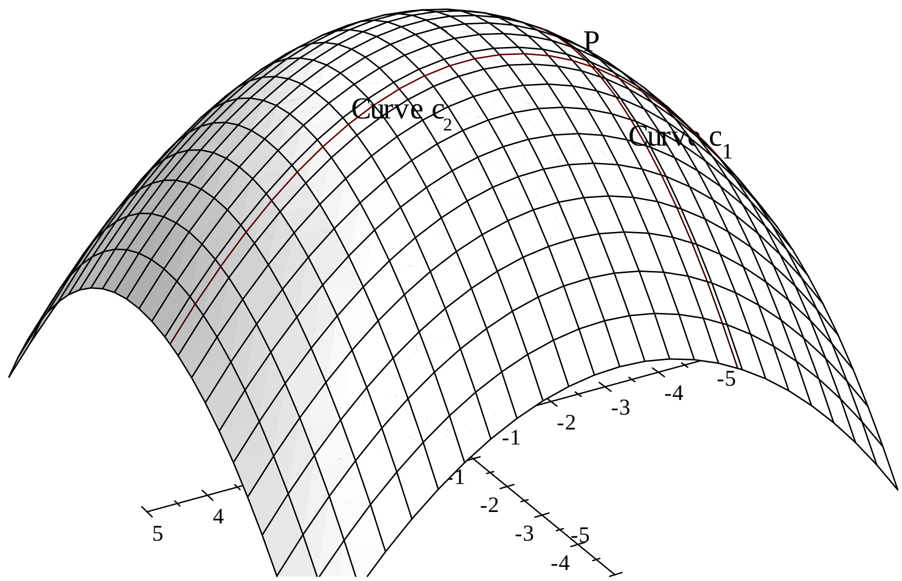


Figure (iv). Curves $\mathbf{c}_1$ and $\mathbf{c}_2$

In particular, the tangent vector $\mathbf{u} =< 1, 0, f_x(c, d) >$ to $\mathbf{c}_1$ and the tangent vector $\mathbf{v} =< 0, 1, f_y(c, d) >$ to $\mathbf{c}_2$ at $(c, d, f(c, d))$ are in the tangent plane. Therefore their cross product

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x(c,d) \\ 0 & 1 & f_y(c,d) \end{vmatrix} = -f_x(c,d)\mathbf{i} - f_y(c,d)\mathbf{j} + \mathbf{k}$$

is a normal to the tangent plane. We now have enough information to determine the equation of the tangent plane at $(c, d, f(c, d))$ and it is given by

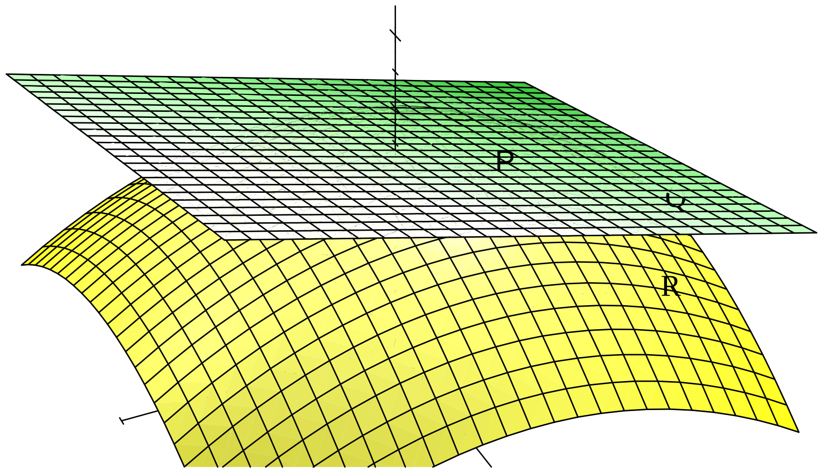
$$(x - c)\,(-f_x(c, d)) + (y - d)\,(-f_y(c, d)) + (z - f(c, d))\,(1) = 0$$

This may be written as

$$z(x, y) = (x - c)\, f_x(c, d) + (y - d)\, f_y(c, d) + f(c, d)$$

When (x, y) is close to (c, d), the point $R\,(x, y, f(x, y))$ in the graph of f is close to the point $Q\,(x, y, z(x, y))$ in the plane, (see the figure below obtained by rotating figure (iii) to show Q and R), therefore

$$f(x, y) \simeq z(x, y) = (x - c)\, f_x(c, d) + (y - d)\, f_y(c, d) + f(c, d). \tag{2}$$



We may rearrange (2) as

$$f(x,y) - f(c,d) \simeq (x-c)\, f_x(c,d) + (y-d)\, f_y(c,d)$$

For convenience, write $(x,y) = (c+h, d+k)$ where $h = x-c$ and $k = y-d$. Then

$$f(c+h, d+k) - f(c,d) = f_x(c,d)h + f_y(c,d)k$$

Thus the difference $f(c+h,d+k) - f(c,d)$ can be approximated by the linear map $L(h,k)$ with formula

$$L(h,k) = f_x(c,d)h + f_y(c,d)k \tag{3}$$

In order for this observation to be a generalization of the differentiability of a function of one variable, we should demand that the error $f(c+h,d+k) - f(c,d) - L(h,k)$ is small compared to $||(h,k)||$ when (h,k) is close to $(0,0)$. This suggests the following definition:

Definition 7 *Let $f : A \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a given function and $(c,d) \in A$. We say that f is differentiable at (c,d) if there is a linear map $L : \mathbb{R}^2 \to \mathbb{R}$ such that*

$$\lim_{(h,k)\to(0,0)} \frac{|f(c+h,d+k) - f(c,d) - L(h,k)|}{||(h,k)||} = 0.$$

L is is called the derivative of f at (c,d) and it is denoted by $Df(c,d)$.

Example 8 *Let $f(x,y) = x^2 + 3xy + y^2$ and $(c,d) = (1,2)$. Take a point $(c+h,d+k) = (1+h,2+k)$. It turns out that*

$$f(1+h, 2+k) - f(1,2) = 8h + 7k + h^2 + 3hk + k^2.$$

The linear term in $(8h+7k) + \left(h^2 + 3hk + k^2\right)$ is $8h+7k$. We have to verify that when (h,k) is close to $(0,0)$, the error term $h^2 + 3hk + k^2$ is small compared to $||(h,k)||$. By the triangle inequality,

$$\left|h^2 + 3hk + k^2\right| \leq h^2 + 3\,(|h|)\,(|k|) + k^2$$

To go beyond this, simply note that $|h| \leq \sqrt{h^2+k^2}$ and $|k| \leq \sqrt{h^2+k^2}$. Therefore

$$\left|h^2 + 3hk + k^2\right| \leq h^2 + 3\left(\sqrt{h^2+k^2}\right)\left(\sqrt{h^2+k^2}\right) + k^2 = 4\left(h^2+k^2\right) = 4\,||(h,k)||^2$$

This implies that

$$\lim_{(h,k)\to(0,0)} \frac{\left|f(c+h,d+k) - f(c,d) - L_{(1,2)}(h,k)\right|}{||(h,k)||} \leq \lim_{(h,k)\to(0,0)} \frac{4\,||(h,k)||^2}{||(h,k)||} = 0.$$

Therefore f is differentiable at $(1,2)$ and $Df(1,2)$ has formula

$$Df(1,2)(h,k) = 8h + 7k$$

In matrix form

$$Df(1,2)(h,k) = \begin{pmatrix} 8 & 7 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = 8h + 7k$$

In general, take an arbitrary point (c,d) and consider the expression $f(c+h, d+k) - f(c,d)$. It simplifies as

$$f(c+h, d+k) - f(c,d) = (2c+3d)\,h + (3c+2d)\,k + 3hk + h^2 + k^2 \tag{4}$$

The linear term in the right hand side of (4) is $(2c+3d)\,h + (3c+2d)\,k$. The error term is $3hk + h^2 + k^2$ and we have already shown that when (h,k) is close to $(0,0)$, it is small compared to $||(h,k)||$, therefore f is differentiable at (c,d) with derivative

$$Df(c,k)\,(h,k) = (2c+3d)\,h + (3c+2d)\,k$$

In matrix form

$$Df(c,k)(h,k) = \begin{pmatrix} 2c+3d & 3c+2d \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = (2c+3d)\,h + (3c+2d)\,k$$

As expected, $2c + 3d = f_x(c,d)$ and $3c + 2d = f_y(c,d)$.

Exercise 9 *For each function f, (i) Evaluate $f(c+h, d+k) - f(c,d)$; (ii) Determine $Df(c,k)(h,k)$; and (iii) Verify that* $\lim\limits_{(h,k)\to(0,0)} \dfrac{|f(c+h,d+k) - f(c,d) - Df(c,k)(h,k)|}{||(h,k)||} = 0.$

(a) $f(x,y) = xy^2 + 3x^2$ *(b)* $f(x,y) = x^3 + 4xy^2 - y$ *(c)* $f(x,y) = 4x + 3y - 7$