# MA 109 Week 6

Mayukh Mukherjee

Department of Mathematics

January 3, 2022

# Partial Derivatives

As before, $U$ will denote a subset of $\mathbb{R}^2$. Given a function $f : U \to \mathbb{R}$, we can fix one of the variables and view the function $f$ as a function of the other variable alone. We can then take the derivative of this one variable function.

To make things precise, fix $x_2$.
Definition: The partial derivative of $f : U \to \mathbb{R}$ with respect to $x_1$ at the point $(a, b)$ is defined by

$$\frac{\partial f}{\partial x_1}(a,b) := \lim_{x_1 \to a} \frac{f((x_1,b)) - f((a,b))}{x_1 - a}.$$

Similarly, one can define the partial derivative with respect to $x_2$. In this case the variable $x_1$ is fixed and $f$ is regarded only as a function of $x_2$:

$$\frac{\partial f}{\partial x_2}(a,b) := \lim_{x_2 \to b} \frac{f((a,x_2)) - f((a,b))}{x_2 - b}.$$

# Directional Derivatives

The partial derivatives are special cases of the directional derivative. Let $v = (v_1, v_2)$ be a unit vector. Then $v$ specifies a direction in $\mathbb{R}^2$.

Definition: The directional derivative of $f$ in the direction $v$ at a point $x = (x_1, x_2)$ is defined as

$$\nabla_v = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} = \lim_{t \to 0} \frac{f((x_1 + tv_1, x_2 + tv_2)) - f((x_1, x_2))}{t}.$$

It measures the rate of change of the function $f$ along the path $x + tv$

If we take $v = (1, 0)$ in the above definition, we obtain $\partial f / \partial x_1$, while $v = (0, 1)$ yields $\partial f / \partial x_2$.

Consider the function

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = 0 \text{ or if } x_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

It should be clear to you that since this function is constant along the two axes,

$$\frac{\partial f}{\partial x_1}(0,0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial x_2}(0,0) = 0$$

On the other hand, $f(x_1, x_2)$ is not continuous at the origin! Thus, a function may have both partial derivatives (and, in fact, any directional derivative - see the next slide) but still not be continuous. This suggests that for a function of two variables, just requiring that both partial derivatives exist is not a good or useful definition of "differentiability".

Recall again, the following function from Exercise 6.5:

$$\frac{x^2y^2}{x^2y^2 + (x-y)^2} \quad \text{for} \quad (x,y) \neq (0,0).$$

Let us further set $f(0,0) = 0$. You can check that every directional derivative exists and is equal to 0, except along $y = x$ when the directional derivative is not defined. However, we have already seen that the function is not continuous at the origin since we have shown that $\lim_{(x,y)\to 0} f(x,y)$ does not exist. For an example with directional derivatives in all directions see Exercise 6.3(i).

Conclusion: All directional derivatives may exist at a point even if the function is discontinuous.

Let us go back and examine the notion of differentiability for a function of $f(x)$ of one variable. Suppose $f$ is differentiable at the point $x_0$, What is the equation of the tangent line through $(x_0, f(x_0))$?

$$y = f(x_0) + f'(x_0)(x - x_0)$$

as the equation for the tangent line. If we consider the difference $f(x) - f(x_0) - f'(x_0)(x - x_0)$ we get the distance of a point on the tangent line from the curve $y = f(x)$. Writing $h = (x - x_0)$, we see that the difference can be rewritten

$$f(x_0 + h) - f(x_0) - f'(x_0)h$$

The tangent line is close to the function $f$ - how close?- so close that even after dividing by $h$ the distance goes to 0. A few lectures ago we wrote this as

$$|f(x_0 + h) - f(x_0) - f'(x_0)h| = p(h)|h|$$

where $p(h)$ is a function that goes to 0 as $h$ goes to 0.

The preceding idea generalises to two (or more) dimensions. Let $f(x, y)$ be a function which has both partial derivatives. In the two variable case we need to look at the distance between the surface $z = f(x, y)$ and its tangent plane.

Let us first recall how to find the equation of a plane passing through the point $P = (x_0, y_0, z_0)$. It is the graph of the function

$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0).$$

Let us determine the tangent plane to $z = f(x, y)$ passing through a point $P = (x_0, y_0, z_0)$ *on the surface*. In other words, we have to determine the constants $a$ and $b$.

If we fix the $y$ variable and treat $f(x, y)$ only as a function of $x$, we get a curve. Similarly, if we treat $g(x, y)$ as function only of $x$, we obtain a line. The tangent to the curve must be the same as the line passing through $(x_0, y_0, z_0)$, and, in any event, their slopes must be the same. Thus, we must have

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0) = a.$$

Arguing in exactly the same way, but fixing the $x$ variable and varying the $y$ variable we obtain

$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial g}{\partial y}(x_0, y_0) = b.$$

Hence, the equation of the tangent plane to $z = f(x, y)$ at the point $(x_0, y_0)$ is

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

# Differentiability for functions of two variables

We now define differentiability for functions of two variables by imitating the one variable definition, but using the "$p(h)$" version. We let $(x, y) = (x_0, y_0) + (h, k) = (x_0 + h, y_0 + k)$

Definition A function $f : U \to \mathbb{R}$ is said to be differentiable at a point $(x_0, y_0)$ if $\frac{\partial f}{\partial x}(x_0, y_0)$, and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and

$$\lim_{(h,k)\to 0} \frac{\left| f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right|}{\|(h, k)\|} = 0,$$

This is saying that the distance between the tangent plane and the surface is going to zero even after dividing by $\|(h, k)\|$. We could rewrite this as

$$\left| f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right| = p(h, k)\|(h, k)\|$$

where $p(h, k)$ is a function that goes to 0 as $\|(h, k)\| \to 0$.This form of differentiability now looks exactly like the one variable version case.

# The derivative as a linear map

We can rewrite the differentiability criterion once more as follows. We define the $1 \times 2$ matrix

$$Df(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}.$$

A $1 \times 2$ matrix can be multiplied by a column vector (which is $2 \times 1$ matrix) to give a real number. In particular:

$$\begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

that is,

$$Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

The definition of differentiability can thus be reformulated using matrix notation.

Definition: The function $f(x, y)$ is said be differentiable at a point $(x_0, y_0)$ if there exists a matrix denoted $Df((x_0, y_0))$ with the property that

$$f((x_0, y_0) + (h, k)) - f(x_0, y_0) - Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} = p(h, k)\|(h, k)\|,$$

for some function $p(h, k)$ which goes to zero as $(h, k)$ goes to zero. Viewing the derivative as a matrix allows us to view it as a linear map from $\mathbb{R}^2 \to \mathbb{R}$. Given a $1 \times 2$ matrix $A$ and two column vectors $v$ and $w$, we see that

$$A \cdot (v + w) = A \cdot v + A \cdot w \quad \text{and} \quad A \cdot (\lambda v) = \lambda(A \cdot v),$$

for any real number $\lambda$. As we have seen before, functions satisfying the above two properties are called linear functions or linear maps. Thus, the map $v \to A \cdot v$ gives a linear map from $\mathbb{R}^2$ to $\mathbb{R}$.

The matrix $Df(x_0, y_0)$ is called the Derivative matrix of the function $f(x, y)$ at the point $(x_0, y_0)$.

# The Gradient

When viewed as a row vector rather than as a matrix, the Derivative matrix is called the gradient and is denoted $\nabla f(x_0, y_0)$. Thus

$$\nabla f(x_0, y_0) = \left( \frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right).$$

In terms of the coordinate vectors $\mathbf{i}$ and $\mathbf{j}$ the gradient can be written as

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j}.$$

# A criterion for differentiability

Before we state the criterion, we note that with our definition of differentiability, every differentiable function is continuous.

**Theorem**

*Let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ exist and are continuous in a neighbourhood of a point $(x_0, y_0)$ (that is in a region of the plane of the form $\{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$ for some $r > 0$. Then $f$ is differentiable at $(x_0, y_0)$.*

We omit the proof of this theorem. However, we note that a function whose partial derivatives exist and are continuous is said to be continuously differentiable or of class $\mathcal{C}^1$. The theorem says that every $\mathcal{C}^1$ function is differentiable.

# Three variables

For the next few slides, we will assume that $f : U \to \mathbb{R}$ is a function of three variables, that is, $U$ is a subset of $\mathbb{R}^3$. In this case, if we denote the variables by $x$, $y$ and $z$, we get three partial derivatives as follows: we hold two of the variables constant and vary the third. For instance if $y$ and $z$ are kept fixed while $x$ is varied, we get the partial derivative with respect to $x$ at the point $(a, b, c)$:

$$\frac{\partial f}{\partial x}(a, b, c) = \lim_{x \to a} \frac{f(x, b, c) - f(a, b, c)}{x - a}.$$

In a similar way we can define the partial derivatives

$$\frac{\partial f}{\partial y}(a, b, c) \quad \text{and} \quad \frac{\partial f}{\partial z}(a, b, c).$$

Once we have the three partial derivatives we can once again define the gradient of $f$:

$$\nabla f(a, b, c) = \left( \frac{\partial f}{\partial x}(a, b, c), \frac{\partial f}{\partial y}(a, b, c), \frac{\partial f}{\partial z}(a, b, c) \right).$$

# Differentiability in three variables

Exercise 1: Formulate a definition of differentiability for a function of three variables.

Exercise 2: Formulate the analogue of the criterion for differentiability for a function of three variables.

We can also define differentiability for functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ where $m$ and $n$ are any positive integers. We will do this in detail in this course when $m$ and $n$ have the values 1 and 2 and 3.

Finally, the rules for the partial derivatives of sums, differences, products and quotients of functions $f, g : U \to \mathbb{R}$, $(U \subset \mathbb{R}^m,\ m = 2, 3)$ are exactly analogous to those for the derivative of functions of one variable.

# The Chain Rule

We now study the situation where we have composition of functions. We assume that $x, y : I \to \mathbb{R}$ are differentiable functions from some interval (open or closed) to $\mathbb{R}$. Thus the pair $(x(t), y(t))$ defines a function from $I$ to $\mathbb{R}^2$. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is differentiable. We would like to study the derivative of the composite function $z(t) = f(x(t), y(t))$ from $I$ to $\mathbb{R}$.

**Theorem**

*With notation as above*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

For a function $w = f(x, y, z)$ in three variables the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

# Clarifications on the notation

The form in which I have written the chain rule is the standard one used in many books (both in engineering and mathematics). However, it is not very good notation. For instance, in the formula

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

the letter $z$ is being used for two different functions: both as a function $z(t)$ from $\mathbb{R}$ to $\mathbb{R}$ on the left hand side, and as a function $z(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$. If one wants to be precise one should write the chain rule as

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Similarly, for the function $w$ we should write

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

# Verifying the chain rule in a simple case

Example: Let us verify this rule in a simple case. Let $z = xy$, $x = t^3$ and $y = t^2$.

Then $z = t^5$ so $z'(t) = 5t^4$. On the other hand, using the chain rule we get

$$z'(t) = y \cdot 3t^2 + x \cdot 2t = 3t^4 + 2t^4 = 5t^4.$$

Example: A continuous mapping $c : I \to \mathbb{R}^n$ of an interval $I$ to $\mathbb{R}$ is called a path or curve in $\mathbb{R}^n$, $(n = 2, 3)$.

In what follows, we will assume that all the curves we have are actually differentiable, not just continuous. Saying that $c(t)$ is a differentiable function of $t$, means that each of $g(t), h(t), k(t)$ are differentiable functions from $\mathbb{R} \to \mathbb{R}$.

# An application to tangents of curves

Let us consider a curve $c(t)$ in $\mathbb{R}^3$. Each point on the curve will be given by a triple of coordinates which will depend on $t$. That is, the curve can be described by a triple of functions $(g(t), h(t), k(t))$.

If we write

$$c(t) = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}, \quad \text{then} \quad c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

represents its tangent or velocity vector at the point $c(t_0)$.

# Tangents to curves on surfaces

So far our example has nothing to do with the chain rule. Suppose $z = f(x, y)$ is a surface, and $c(t) = (g(t), h(t), f(g(t), h(t))$ lies on the $z = f(x, y)$ (here we are assuming that $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function!) Let us compute the tangent vector to the curve at $c(t_0)$. It is given by

$$c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

where $k(t) = (f(g(t), h(t))$. Using the chain rule we see that

$$k'(t_0) = \frac{\partial f}{\partial x} g'(t_0) + \frac{\partial f}{\partial y} h'(t_0).$$

We can further show that this tangent vector lies on the tangent plane to the surface $z = f(x, y)$. Indeed we have already seen that the tangent plane has the equation

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

A normal vector to this plane is given by

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1\right).$$

Thus, to verify that the tangent vector lies on the plane, we need only check that its dot product with normal vector is 0. But this is now clear.

Just to give a concrete example of what we are talking about, take a curve $(g(t), h(t))$ in the unit disc $x^2 + y^2 \leq 1$ in the $xy$ plane. Then $\left(g(t), h(t), \sqrt{1 - g(t)^2 - h(t)^2}\right)$ lies on the upper hemisphere $z = \sqrt{1 - x^2 - y^2}$. For concreteness, we can take $I = \left[0, \frac{1}{\sqrt{2}}\right]$, $g(t) = t$ and $h(t) = t^2$.

# Another application: Directional derivatives

Let $U \subset \mathbb{R}^3$ and let $f : U \to \mathbb{R}$ be differentiable. We want to relate the directional derivative to the gradient,

We consider the (differentiable) curve $c(t) = (x_0, y_0, z_0) + tv$, where $v = (v_1, v_2, v_3)$ is a unit vector. We can rewrite $c(t)$ as $c(t) = (x_0 + tv_1, y_0 + tv_2, z_0 + tv_3)$. We apply the chain rule to compute the derivative of the function $f(c(t))$:

$$\frac{df}{dt} = \frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 + \frac{\partial f}{\partial z} v_3.$$

But the left hand side is nothing but the directional derivative in the direction $v$. Hence,

$$\nabla_v f = \frac{df}{dt} = \nabla f \cdot v.$$

Of course, the same argument works when $U \subset \mathbb{R}^2$ and $f$ is a function of two variables.

# The Chain Rule and Gradients

The preceding argument is a special case of a more general fact. Let $c(t)$ be any curve in $\mathbb{R}^3$. Then, clearly by the chain rule we have

$$\frac{df}{dt} = \nabla f(c(t)) \cdot c'(t).$$

I leave this to you as a simple exercise.

Going back to the directional derivative, we can ask ourselves the following question. In what direction is $f$ changing fastest at a given point $(x_0, y_0, z_0)$? In other words, in which direction does the directional derivative attain its largest value?

Using what we have just learnt, we are looking for a unit vector $v = (v_1, v_2, v_3)$ such that

$$\nabla f(x_0, y_0, z_0) \cdot v$$

is as large as possible.

We rewrite the preceding dot product as

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\|\|v\| \cos\theta.$$

where $\theta$ is the angle between $v$ and $\nabla f(x_0, y_0, z_0)$.
Since $v$ is a unit vector this gives

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \cos\theta.$$

The maximum value on the right hand side is obviously attained when $\theta = 0$, that is, when $v$ points in the direction of $\nabla f$. In other words the function is increasing fastest in the direction $v$ given by $\nabla f$. Thus the unit vector that we seek is

$$v = \frac{\nabla f(x_0, y_0, z_0)}{\|\nabla f(x_0, y_0, z_0)\|}.$$

# Surfaces defined implicitly

So far we have only been considering surfaces of the form $z = f(x, y)$, where $f$ was a function on a subset of $\mathbb{R}^2$. We now consider a more general type of surface $S$ defined implicitly:

$$S = \{(x, y, z) \mid f(x, y, z) = b\},$$

where $b$ is a constant. Most surfaces we have come across are usually described in this form, for instance, the sphere which is given by $x^2 + y^2 + z^2 = r^2$ or the right circular cone $x^2 + y^2 - z^2 = 0$. Let us try to understand what a tangent plane is more precisely.

If $S$ is a surface, a tangent plane to $S$ at a point $s \in S$ (if it exists) is a plane that contains the tangent lines at $s$ to all curves passing through $s$ and lying on $S$.

For instance, with the definition above, it is clear that a tangent plane to the right circular cone does not exist at the origin, since such a plane would have to contain the lines $x = 0, y = z$, $x = 0, y = -z$ and $y = 0, x = z$. Clearly no such plane exists.

If $c(t)$ is an curve on the surface $S$ given by $f(x, y, z) = b$, we see that

$$\frac{d}{dt} f(c(t)) = 0.$$

On the other hand, by the chain rule,

$$0 = \frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

Thus, if $s = c(t_0)$ is a point on the surface, we see that

$$\nabla f(c(t_0)) \cdot c'(t_0) = 0,$$

for every curve $c(t)$ on the surface $S$ passing through $t_0$. Hence, if $\nabla f(c(t_0)) \neq 0$, then $\nabla f(c(t_0))$ is perpendicular to the tangent plane of $S$ at $s_0$.

Let $\mathbf{r}$ denote the position vector

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

of a point $P = (x, y, z)$ in $\mathbb{R}^3$. Instead of writing $\|\mathbf{r}\|$, it is customary to write $r$. This notation is very useful. For instance, Newton's Law of Gravitation can be expressed as

$$\mathbf{F} = -\frac{GMm}{r^3} \cdot \mathbf{r},$$

where the mass $M$ is assumed to be at the origin, $\mathbf{r}$ denotes the position vector of the mass $m$, $G$ is a constant and $\mathbf{F}$ denotes the gravitational force between the two (point) masses.

A simple computation shows that

$$\nabla\left(\frac{1}{r}\right) = -\frac{\mathbf{r}}{r^3}.$$

Thus the gravitational force at any point can be expressed as the gradient of a function. Moreover, it is clear that

$$\left\|\nabla\left(\frac{1}{r}\right)\right\| = \left\|-\frac{\mathbf{r}}{r^3}\right\| = \frac{1}{r^2}.$$

Keeping our previous discussion in mind, we know that if $V = GMm/r$, $\mathbf{F} = \nabla V$.

What are the level surfaces of $V$? Clearly, $r$ must be a constant on these level sets, so the level surfaces are spheres. Since $\mathbf{F}$ is a multiple of $-\mathbf{r}$, we see that $F$ points towards the origin and is thus orthogonal to the sphere.

In order to make our notation less cumbersome, we introduce the notation $f_x$ for the partial derivative $\frac{\partial f}{\partial x}$. The notations $f_y$ and $f_z$ will have the obvious meanings.

Since we know that the gradient of $f$ is normal to the level surface $S$ given by $f(x, y, z) = c$ (provided the gradient is non zero), it allows us to write down the equation of the tangent plane of $S$ at the point $s = (x_0, y_0, z_0)$. The equation of this plane is

$$f_x(x_0, y_0, z_0)(x - x_0) + f_y(x_0, y_0, z_0)(y - y_0) + f_z(x_0, y_0, z_0)(z - z_0) = 0.$$

For the curve $f(x, y) = c$ we can similarly write down the equation of the tangent passing through $(x_0, y_0)$:

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$$

Note that the fact that the gradient of $f$ is normal to the level surface $f(x, y, z) = c$ is true only for implicitly defined surfaces. If the surface is given as $z = f(x, y)$, then we cannot simply take the gradient of $f$ and make the same statement. We must first convert our explicit surface to the implicit surface $S$ given by $g(x, y, z) = z - f(x, y) = 0$. Then $\nabla g$ will be normal to $S$.

# The proof of the chain rule

How does one actually prove the chain rule for a function $f(x, y)$ of two variables? We can write

$$f(x(t+h), y(t+h)) = f(x(t) + h[x'(t) + p_1(h)], y(t) + h[y'(t) + p_2(h)])$$

for functions $p_1$ and $p_2$ that go to zero as $h$ goes to zero. Here we are simply using the differentiability of $x$ and $y$ as functions of $t$. Now we can write the right hand side as

$$f(x(t), y(t)) + Df(h[x'(t) + p_1(h)], h[y'(t) + p_2(h)]) + p_3(h)h$$

by using the differentiability of $f$, for some other function $p_3(h)$ which goes to zero as $h$ goes to zero (you may need to think about this step a little). This gives

$$f(x(t+h), y(t+h)) - f(x(t), y(t)) - f_x x'(t)h - f_y y'(t)h = p(h)h.$$

# Functions from $\mathbb{R}^m \to \mathbb{R}^n$

So far we have only studied functions whose range was a subset of $\mathbb{R}$. Let us now allow the range to be $\mathbb{R}^n$, $n = 1, 2, 3, \ldots$. Can we understand what continuity, differentiability etc. mean?

Let $U$ be a subset of $\mathbb{R}^m$ ($m = 1, 2, 3, \ldots$) and let $f : U \to \mathbb{R}^n$ be a function. If $x = (x_1, x_2, \ldots, x_m) \in U$, $f(x)$ will be an $n$-tuple where each coordinate is a function of $x$. Thus, we can write $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, where each $f_i(x)$ is a function from $U$ to $\mathbb{R}$.

Functions which take values in $\mathbb{R}$ are called scalar valued functions, which functions which take values in $\mathbb{R}^n$, $n > 1$ are usually called vector valued functions.

# Vector fields

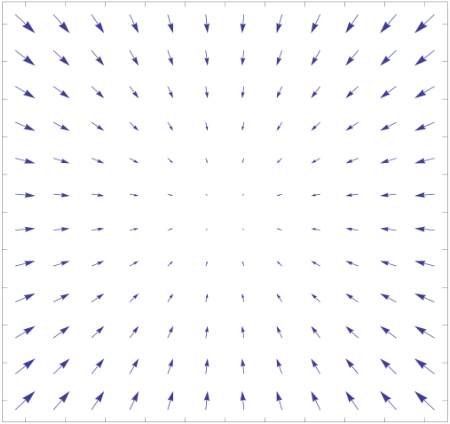
When $m = n$, vector valued functions are often called vector fields. We will study vector fields in slightly greater detail when $m = n = 2$ and $m = n = 3$.

We have already seen one example of a vector field - the gravitational force field $-\frac{GMm}{r^3} \cdot \mathbf{r}$ felt by a mass $m$ whose position vector with respect to a mass $M$ at the origin is $\mathbf{r}$. In this particular case we showed the the force field arose as the gradient of a scalar valued function (the potential $V = GMm/r$).

One of the most important questions in calculus is the following: Given a vector field, when does it arise as the gradient of a scalar function?. In physics, vector force fields that arise from a scalar potential function are called Conservative.

# Some pictures of vector fields

We can actually visualize two dimensional vector fields as follows. At each point in $\mathbb{R}^2$ we can draw an arrow starting at that point pointing in the direction of the image vector and with size proportional to the magnitude of the image vector.



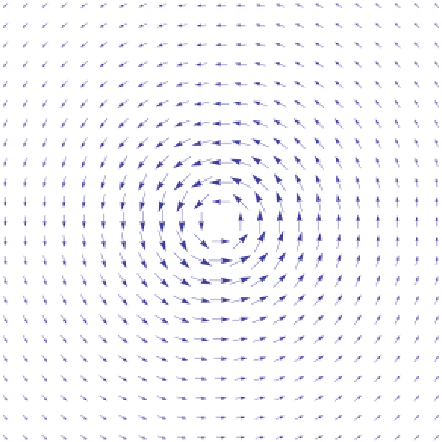
What function from $\mathbb{R}^2$ to $\mathbb{R}^2$ does this picture represent?

$$f(x, y) = (-x, -y)$$

the the radial vector field.

http://en.wikipedia.org/wiki/File:Radial_vector_field_dense.svg

How about this one?

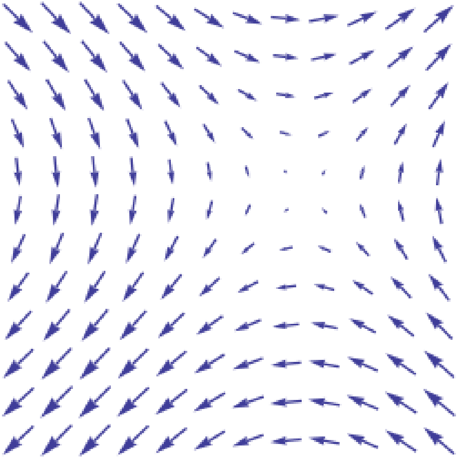


$$f(x, y) = (-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2})$$

This is an example of an irrotational vector field.

It cannot be written as the gradient of a potential function.

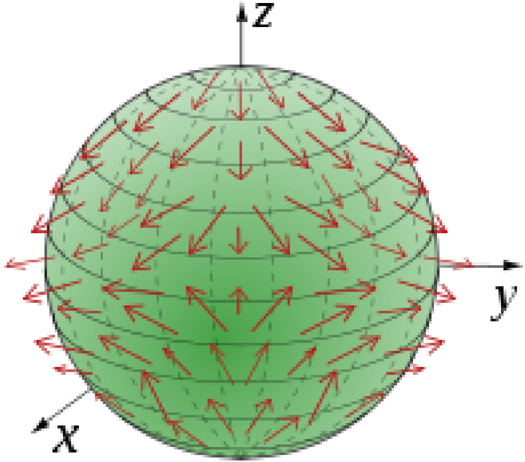
Here is another (more complicated one)



http://en.wikipedia.org/wiki/File:VectorField.svg

$$f(x, y) = (\sin y, \sin x)$$

One can also talk about two dimensional vector fields on any two dimensional surface. Here is a picture of a vector field on a sphere.




http://en.wikipedia.org/wiki/File:Vector_sphere.svg

# Vector fields in the real world

Many real world phenomena can be understood using the language of vector fields. In physics, apart from gravitation, electromagnetic forces can also be represented by vector fields. That is, to each point in space we attach the vector representing the force at that point. Such fields are called force fields.

Fluids flowing are also often modeled using vector fields, with each point being mapped to the vector representing the velocity of the fluid flow. For instance, the velocity of winds in the atmosphere can be represented as a vector field. Such fields are called velocity fields.

# Higher derivatives

Just as we repeatedly differentiated a function of one variable to get higher derivatives, we can also look at higher partial derivatives.

However, we now have more freedom. If we have a function $f(x_1, x_2)$ of two variables, we could first take the partial derivative with respect to $x_1$, then with respect to $x_2$, then again with respect to $x_2$, and so on. Does the order in which we differentiate matter?

**Theorem**

*Let $f : U \to R$ be a function such that the partial derivatives $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right)$ exist and are continuous for every $1 \leq i, j \leq m$. Then,*

$$\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i}(f)\right).$$

Functions $f : U \to \mathbb{R}$ for which the mixed partial derivatives of order 2 (that is, the $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}(f)\right)$) are all continuous are called $\mathcal{C}^2$ functions. The theorem above says that for $\mathcal{C}^2$ functions, the order in which one takes partial derivatives does not matter.

From now on we will use use the following notation. By

$$\frac{\partial^n f}{\partial x_1^{n_1} \partial x_2^{n_2} \dots \partial x_k^{n_k}},$$

we mean: first take the partial derivative of $f$ $n_1$ times with respect to $x_1$, then $n_2$ times with respect to $x_2$, and so on. The number $n$ is nothing but $n_1 + n_2 + \ldots + n_k$. It is called the order of the mixed partial derivative.

Finally, we say that a function is $\mathcal{C}^k$ if all mixed partial derivatives of order $k$ exist and are continuous. A function $f : U \to \mathbb{R}^n$ will be said to be $\mathcal{C}^k$ if each of the functions $f_1, f_2, \ldots, f_n$ are $\mathcal{C}^k$ functions.

From the preceding slide we see the we can talk about $\mathcal{C}^k$ functions for any function from (a subset of) $\mathbb{R}^m$ to $\mathbb{R}^n$. As in the one variable case we can also talk of smooth functions - these are functions for which all partial derivatives of all orders exist. In particular, the notion of a smooth vector field makes perfect sense.

There are many interesting facts about smooth vector fields. I will mention just one:

You cannot comb a porcupine.

Or, in more mathematical terms, every smooth vector field on the sphere will vanish at at least one point.

Note that we require at each point on the sphere, the vector we assign must lie in the plane tangent to the sphere at that point.

# Local maxima and minima

As in the one variable case we can define local maxima and minima for a function of two or more variables. These definitions can be made for any function. They do not require us to assume any differentiability properties for the functions. Let $f : U(\subset \mathbb{R}^2) \to \mathbb{R}$ be a function of two variables.

Definition: We will say that the function $f(x, y)$ attains a local minimum at the point $(x_0, y_0)$ (or that $(x_0, y_0)$ is a local minimum point of $f$) if there is a disc

$$D_r(x_0, y_0) = \{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$$

of radius $r > 0$ around $(x_0, y_0)$ such that $f(x, y) \geq f(x_0, y_0)$ for every point $(x, y)$ in $D_r(x_0, y_0)$.

Similarly, we can define a local maximum point (Do this).

# Critical Points

When the function is differentiable we can use the properties of the partial derivatives to find local maxima and minima. As in the one variable situation, we have the first derivative test. This is the analogue of Fermat's theorem. Before formulating the test we need make the following definition.

Defintion: A point $(x_0, y_0)$ is called a critical point of $f(x, y)$ if

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0.$$

What does this say in geometric terms? Recall that the tangent plane to $z = f(x, y)$ at $(x_0, y_0)$ is given by

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Hence, at a critical point, the tangent plane is horizontal, that is, it is parallel to the $xy$-plane.

# The first derivative test

**Theorem**

*If $(x_0, y_0)$ is a local extremum point (that is, a minimum or a maximum point) and if $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist, then $(x_0, y_0)$ is a critical point.*

The proof is similar in the one variable case. If $(x_0, y_0)$ is not a critical point, then at least one of the two partial derivatives must be non-zero. Without loss of generality we can assume that $f_x(x_0, y_0) \neq 0$.

Suppose $f_x(x_0, y_0) > 0$. This means that

$$\lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} > 0.$$

This means that for $|h|$ small enough,

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} > 0.$$

If $h > 0$, this shows that the numerator is positive. On the other hand, if $h < 0$, the numerator must be negative.

Thus, in any disc $D_r(x_0, y_0)$ there are points $(x, y)$ for which $f(x_0, y_0) < f(x, y)$ and $f(x_0, y_0) > f(x, y)$. The same argument can be repeated if $f_x(x_0, y_0) < 0$, giving a contradiction to the fact that $(x_0, y_0)$ is an extreme point . $\square$

# Towards a second derivative test

As in the one variable case, we would like to decide whether a local extremum is a local maximum or a local minimum. In order to this we will need to look a the partial derivatives of order 2. Let us assume that these exist.

We start by defining the Hessian of $f$. This is the matrix

$$\begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix}.$$

From now on we will assume that $f$ is a $\mathcal{C}^2$ function. Recall that this means that $f_{xy} = f_{yx}$.

The determinant of the Hessian is sometimes called the discriminant and is sometimes denoted $D$. Explicitly,

$$D = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - [f_{xy}(x_0, y_0)]^2.$$

# The second derivative test

We give a test for finding local maxima and minima below. In the two variable situation, we will also need to understand what a saddle point is. We will explain this after stating the theorem.

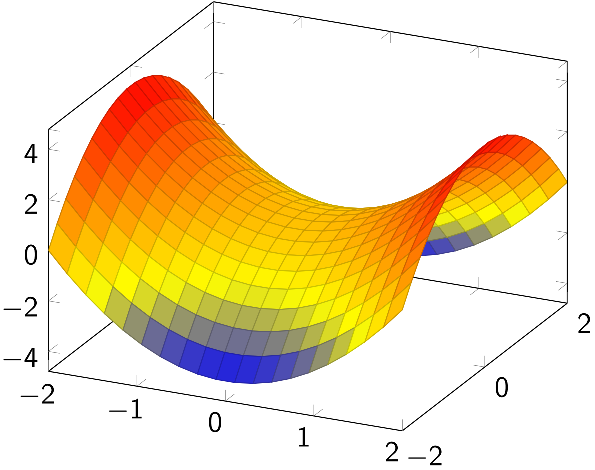
**Theorem**

*With notation as above:*

1. *If* $D > 0$ *and* $f_{xx}(x_0, y_0) > 0$, *then* $(x_0, y_0)$ *is a local minimum for* $f$.
2. *If* $D > 0$ *and* $f_{xx}(x_0, y_0) < 0$, *then* $(x_0, y_0)$ *is a local maximum for* $f$.
3. *If* $D < 0$, *then* $(x_0, y_0)$ *is a saddle point for* $f$.
4. *If* $D = 0$, *further examination of the function is necessary.*

# Saddle points

Since a picture is worth a thousand words, let us start with one.




The point $(0,0)$ is called a saddle point. This is a picture of the graph of $z = x^2 - y^2$.

# An example (from Marsden and Tromba)

Example 1: Find the maxima, minima and saddle points of $z = (x^2 - y^2)e^{\frac{(-x^2-y^2)}{2}}$.

Solution: Let us first find the critical points:

$$\frac{\partial z}{\partial x} = [2x - x(x^2 - y^2)]e^{\frac{(-x^2-y^2)}{2}} \quad \text{and}$$

$$\frac{\partial z}{\partial y} = [-2y - y(x^2 - y^2)]e^{\frac{(-x^2-y^2)}{2}}.$$

Hence the critical points are the simultaneous solutions of

$$x[2 - (x^2 - y^2)] = 0 \quad \text{and} \quad y[-2 - (x^2 - y^2)] = 0$$

The critical points thus lie at

$$(0, 0), \quad (\pm\sqrt{2}, 0), \quad \text{and} \quad (0, \pm\sqrt{2})$$

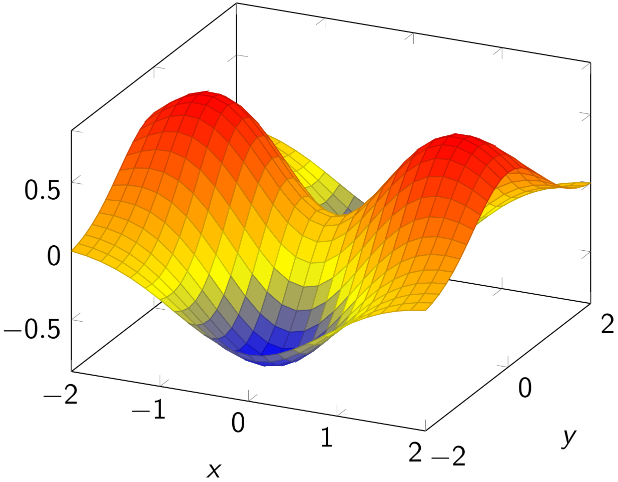
Next we have to find the partial derivatives of order 2. We have

$$\frac{\partial^2 z}{\partial x^2} = [2 - 5x^2 + x^2(x^2 - y^2) + y^2]e^{\frac{(-x^2-y^2)}{2}},$$
$$\frac{\partial^2 z}{\partial x \partial y} = xy(x^2 - y^2)e^{\frac{(-x^2-y^2)}{2}} \quad \text{and}$$
$$\frac{\partial^2 z}{\partial y^2} = [5y^2 - 2 + y^2(x^2 - y^2) - x^2]e^{\frac{(-x^2-y^2)}{2}}.$$

Using the second derivative test we obtain the following table:

| Point | $f_{xx}$ | $f_{xy}$ | $f_{yy}$ | D | Type |
|---|---|---|---|---|---|
| $(0,0)$ | $2$ | $0$ | $-2$ | $-4$ | Saddle |
| $(\sqrt{2},0)$ | $-4/e$ | $0$ | $-4/e$ | $16/e^2$ | Maximum |
| $(-\sqrt{2},0)$ | $-4/e$ | $0$ | $-4/e$ | $16/e^2$ | Maximum |
| $(0,\sqrt{2})$ | $4/e$ | $0$ | $4/e$ | $16/e^2$ | Minimum |
| $(0,-\sqrt{2})$ | $4/e$ | $0$ | $4/e$ | $16/e^2$ | Minimum |

# The previous example in a picture




This is the graph of $z = (x^2 - y^2)e^{\frac{-x^2-y^2}{2}}$.

# Quadratic functions in two variables

Consider functions of the form

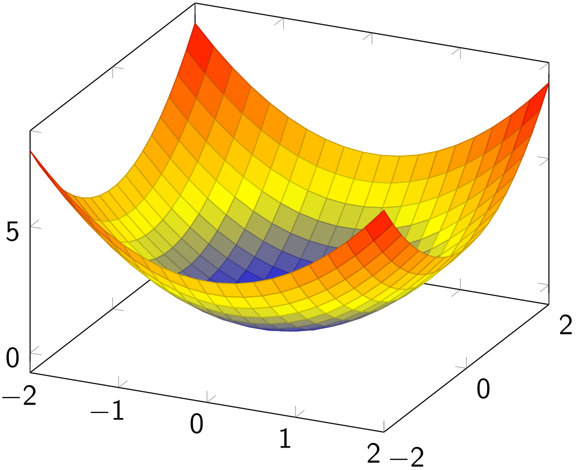
$$z = g(x, y) = Ax^2 + 2Bxy + Cy^2.$$

Notice that $(0, 0)$ is obviously a critical point for the function $g(x, y)$. With a little bit of work we can show that if $AC - B^2 \neq 0$, then $(0, 0)$ is the only critical point of $g$.

From now on we assume that $AC - B^2 \neq 0$. A little more analysis will show the following:

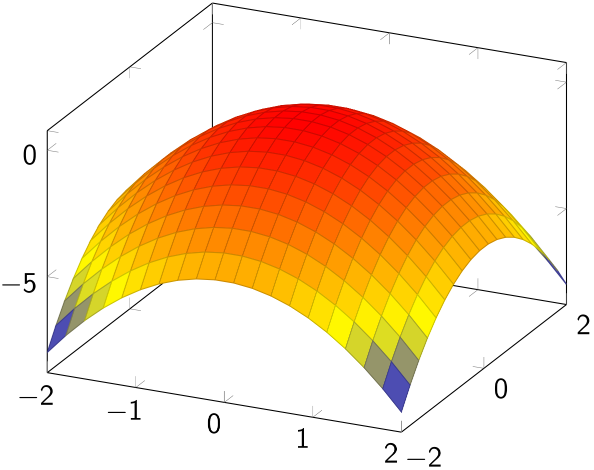
1. If $AC - B^2 > 0$, the function $g$ has a local minimum if $A > 0$ and a local maximum if $A < 0$.
2. If $AC - B^2 < 0$, the function $g$ has a saddle point, that is, in a small disc around the point, the function does not lie on any one side of its tangent plane.

# A local minimum



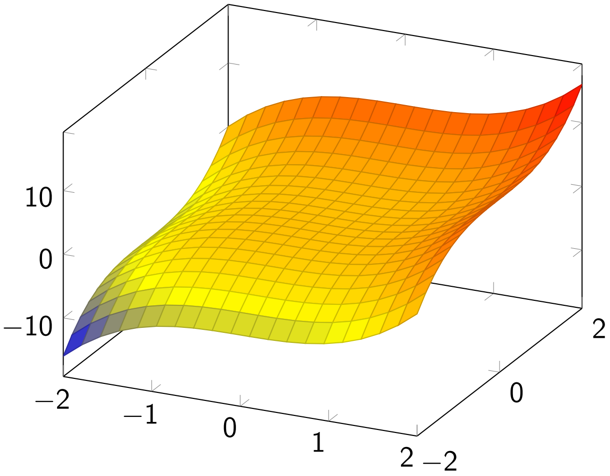

The graph of $x^2 + y^2$ has a local minimum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A > 0$.

# A local maximum



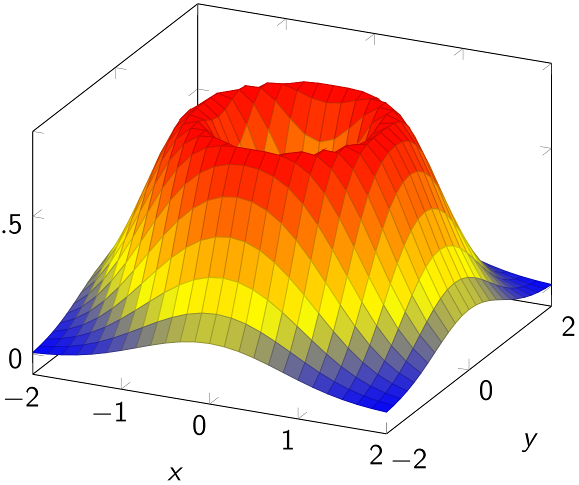

The graph of $-x^2 - y^2$ has a local maximum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A < 0$.

# Where the test is inconclusive




The graph of $x^3 + y^3$. The test is inconclusive at $(0, 0)$.

# The volcano




This is the graph of $z = 2(x^2 + y^2)e^{-x^2-y^2}$. Here the maxima lie on a circle (the rim of the volcano). This sort of behavior cannot arise in a quadratic surface.

# Taylor's theorem in two variables

If we look at the quadratic surface $z = Ax^2 + 2Bxy + Cy^2$, we see that $2A = f_{xx}$, $2B = f_{xy}$ and $2C = f_{yy}$. The second derivative test tells us that whatever is true for quadratic surfaces is true in general. Why is this true? The answer lies in a two variable form of Taylor's Theorem:

**Theorem**

*If $f$ is a $\mathcal{C}^2$ function in a disc around $(x_0, y_0)$, then*

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + f_x h + f_y k$$
$$+ \frac{1}{2!}[f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2] + \tilde{R}_2(h, k),$$

*where $\tilde{R}_2(h, k)/\|(h, k)\|^2 \to 0$ as $\|(h, k)\| \to 0$.*

# From quadratics surfaces to general surfaces

If $(x_0, y_0)$ is a critical point, Taylor's theorem in a disc around the critical point becomes

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + \frac{1}{2!}[f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2] + \tilde{R}_2(h, k),$$

where $\tilde{R}_2(h, k)/\|(h, k)\|^2 \to 0$ as $\|(h, k)\| \to 0$.

Thus, in a small disc around $(x_0, y_0)$ the function $f(x, y)$ looks very much like a quadratic surface, and from the point of view of the critical points there is, in fact, no difference, because the error term can be made as small as we please even after dividing by $\|(h, k)\|^2$. This is why the second derivative test works.

# Back to Taylor's Theorem

Suppose $g : [u, v] \to \mathbb{R}$ is a function of one variable. Let us assume that $g$ is twice continuously differentiable on $[u, v]$. For points $a, b \in (u, v)$ we can rewrite Taylor's Theorem as

$$g(b) = g(a) + g'(a)h + \frac{g''(a)}{2!}h^2 + \frac{g''(c) - g''(a)}{2!}h^2,$$

for some $c$ between $a$ and $b$. Since we have assumed that $g''$ is continuous we see that $(g''(c) - g''(a)) \to 0$ as $h \to 0$. Thus we can write

$$g(b) = g(a) + g'(a)h + \frac{g''(a)}{2!}h^2 + \tilde{R}_2(h),$$

where $\tilde{R}_2(h)/h^2 \to 0$.

Exercise 1: Let $f(x, y)$ be a $\mathcal{C}^2$ function of two variables. Apply the preceding version of Taylor's Theorem to the function

$$g(t) = f(tx + (1 - t)x_0, ty + (1 - t)y_0),$$

for $0 \leq t \leq 1$. This will give the two variable version of Taylor's Theorem stated above. You can easily generalize this to degree $n$.

# Boundedness of continuous functions of two variables

In one variable we saw that continuous functions are bounded in closed bounded intervals. More generally, we can take a finite union of such intervals and the function will remain bounded. Such sets are called compact sets. What is the analogue for $\mathbb{R}^2$?

It is a little harder to define compact sets in $\mathbb{R}^2$, but we can give examples. The closed disc

$$\bar{D}_r = \{(x, y) \mid \|(x, y) - (x_0, y_0)\| \leq r\}$$

of radius $r$ around a point $(x_0, y_0) \in \mathbb{R}^2$ is an example. Another example is the closed rectangle:

$$\bar{S} = \{(x, y) \mid |x - x_0| \leq a, |y - y_0| \leq b\}$$

Finite unions of such sets will also be compact sets. As before, we have

Theorem 32: A continuous function on a compact set in $\mathbb{R}^2$ will attain its extreme values.

# Global extrema as local extrema

Definition: A point $(x_0, y_0)$ such that $f(x,y) \leq f(x_0, y_0)$ or $f(x,y) \geq f(x_0, y_0)$ for all $(x,y)$ in the domain being considered is called a global maximum or minimum point respectively.

In light of Theorem 32, these always exist for continuous functions on closed rectangles or discs.

Assume now that $f$ is a $\mathcal{C}^2$ function on a closed rectangle $\bar{S}$ as above. We can find the global maximum or minimum as follows. We first study all the local extrema which by definition lie in the open rectangle

$$S = \{(x,y) \,|\, |x - x_0| < a, |y - y_0| < b\}$$

(This is because, in order to have a local extremum we need to have a whole disc around the point. A point on the boundary of the closed rectangle does not have such a disc around it where the function is defined.)

After determining all the local maxima we take the points where the function takes the largest value - say $M_1$. We compare this with the maximum value of the function on the boundary of the closed rectangle, say $M_2$. Let $M$ be the maximum of these two values. The points where $M$ is attained are the global maxima.

We can treat a function defined on the closed disc in the same way. Once again, the local extrema will have to lie in the open disc and we will have to consider these values as well as the values of the function on the boundary circle.

# Finding global extrema in $\mathbb{R}^2$

Sometimes, global extrema may exist even when the domain in $\mathbb{R}^2$ (which is not compact, in fact, not even bounded).

For instance, in the example $z = 2(x^2 - y^2)e^{-\frac{(x^2+y^2)}{2}}$, we can check that when $x^2 + y^2 > 16$, $z < 1/2$, that is, outside the disc $\bar{D}_4(0,0)$ (try proving this - I have not chosen this disc optimally).

In the closed disc $\bar{D}_4(0,0)$ we have already found the critical points and the local maxima and minima. We see that $f(\sqrt{2},0) = f(-\sqrt{2},0) = 2/e > 1/2$. There cannot be other local maxima since we have checked all the other critical points, and local maxima can occur only at the critical points. Hence, these points are global maxima in $\bar{D}_4(0,0)$.

Now, outside the disc $\bar{D}_4(0,0)$, we know that $z = f(x,y) < 1/2$. Hence, we see that the value $2/e > 1/2$ is the maximum value taken on all of $\mathbb{R}^2$. Thus, this particular function actually has a global maximum (in fact, two global maxima) on $\mathbb{R}^2$.