MA 109, Week-6

Sanjoy Pusti

Department of Mathematics

December 8, 2022

Functions with range contained in $\mathbb{R}$

We will be interested in studying functions $f : \mathbb{R}^m \to \mathbb{R}$, when $m = 2, 3$. We have already mentioned how limits of such functions can be studied in the first few lectures. Before doing this in detail, however, we will study certain other features of functions in two and three variables.

The most basic thing one needs to understand about a function is the domain on which it is defined. Very often a function is given by a formula which makes sense only on some subset of $\mathbb{R}^m$ and not on the whole of $\mathbb{R}^m$.

When studying functions of two or more variables given by formulae it makes sense to first identify this subset, which is sometimes called the natural domain of the function, and to describe it geometrically if possible.

Exercise 5.1: Find the natural domains of the following functions:
(i) $\frac{xy}{x^2-y^2}$

Clearly this function is defined whenever the denominator is not zero, in other words when $x^2 - y^2 \neq 0$.

The natural domain is thus

$$\mathbb{R}^2 \setminus \{(x, y) \mid x^2 - y^2 = 0\},$$

that is, $\mathbb{R}^2$ minus the pair of straight lines with slopes ± 1.

(ii) $f(x, y) = \log(x^2 + y^2)$

This function is defined whenever $x^2 + y^2 \neq 0$, in other words, in $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Level curves and contour lines

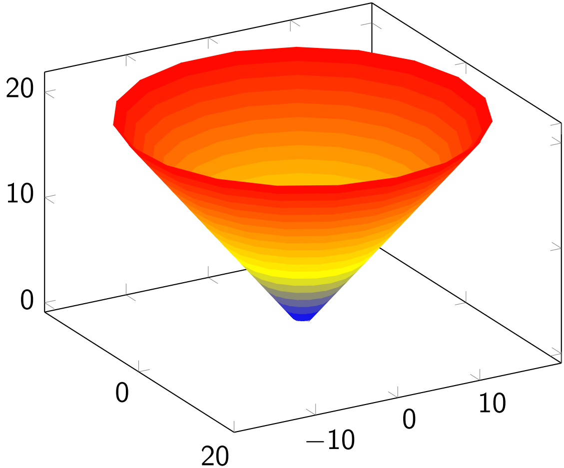
The second thing one should do with a function from $\mathbb{R}^2 \to \mathbb{R}$ is to study its range. This is done in different ways.

One way is to study the level sets of the functions. These are the sets of the form $f(x, y) = c$, where c is a constant. The level set "lives" in the xy-plane.

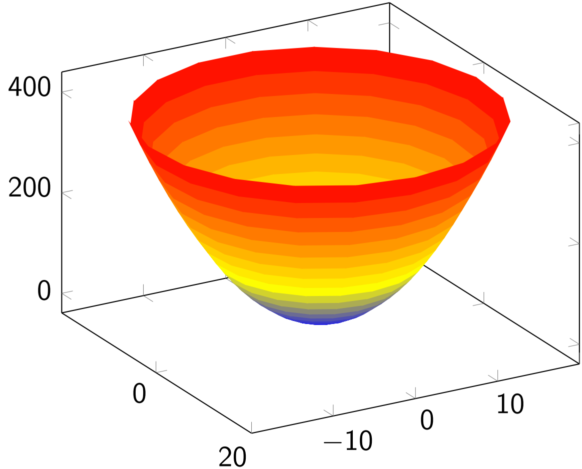
One can also plot (in three dimensions) the surface $z = f(x, y)$. By varying the value of c in the level curves one can get a good idea of what the surface looks like.

When one plots the $f(x, y) = c$ for some constant c one gets a curve. Such a curve is usually called a contour line (the contour "lives" in the $z = c$ plane).

I have a couple of pictures in the next two slides to illustrate the point.



This is the graph of the function $z = \sqrt{x^2 + y^2}$ lying above the xy-plane. It is a right circular cone.
The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves given by $z = f(x, 0)$ and $z = f(0, y)$ give pairs of straight lines in the planes $y = 0$ and $x = 0$.



This is the graph of the function $z = x^2 + y^2$ lying above the xy-plane. It is a paraboloid of revolution.

The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves $z = f(x, 0)$ or $z = f(y, 0)$ give parabolæ lying in the planes $y = 0$ and $x = 0$. Exercise 5.2.(ii).

Limits

We have already said what it means for a function of two or more variables to approach a limit. We simply have to replace the absolute value function on $\mathbb{R}$ by the distance function on $\mathbb{R}^m$. We will do this in two variables. The three variable definition is entirely analogous. We will denote by U a set in $\mathbb{R}^2$.

Definition: A function $f : U \to \mathbb{R}$ is said to tend to a limit l as $x = (x_1, x_2)$ approaches $c = (c_1, c_2)$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x) - l| < \epsilon,$$

whenever $0 < \|x - c\| < \delta$.

We recall that

$$\|x\| = \sqrt{x_1^2 + x_2^2}.$$

Continuity

Before talking about continuity we remark the following. In the plane $\mathbb{R}^2$ it is possible to approach the point c from infinitely many different directions - not just from the right and from the left. In fact, one may not even be approaching the point c along a straight line! Hence, to say that a function from $\mathbb{R}^2$ to $\mathbb{R}$ possesses a limit is actually imposing a strong condition - for instance, the limits along all possible curves leading to the point must exist and all these (infinitely many) limits must be equal.

Once we have the notion of a limit, the definition of continuity is just the same as for functions of one variable.

Definition: The function $f : U \to \mathbb{R}$ is said to be continuous at a point c if

$$\lim_{x \to c} f(x) = f(c).$$

The rules for limits and continuity

The rules for addition, subtraction, multiplication and division of limits remain valid for functions of two variables (or three variables for that matter). Nothing really changes in the statements or the proofs.

Using these rules, we can conclude, as before, that the sum, difference, product and quotient of continuous functions are continuous (as usual we must assume that the denominator of the quotient is non zero).

Continuity through examples

Once again, we emphasise that continuity at a point c is a very powerful condition (since the existence of a limit is implicit).

Exercise 5.3.(i) asks whether the function

$$f(x,y) = \begin{cases} \frac{x^3y}{x^6+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is continuous at $(0,0)$.

Solution: Let us look at the sequence of points $z_n = (\frac{1}{n}, \frac{1}{n^3})$, which goes to $(0,0)$ as $n \to \infty$. Clearly $f(z_n) = \frac{1}{2}$ for all n, so

$$\lim_{n\to\infty} f(z_n) = \frac{1}{2} \neq 0.$$

This shows that f is not continuous at 0.

But does the limit exist?

Iterated limits

When evaluating a limit of the form $\lim_{(x,y)\to(a,b)} f(x,y)$ one may naturally be tempted to let x go to a first, and then let y go to b. Does this give the limit in the previous sense?

Theorem: Let $\lim_{(x,y)\to(a,b)} f(x,y) = L$. Suppose that

$$\lim_{x\to a} f(x,y) = g(y) \text{ and } \lim_{y\to b} f(x,y) = h(x).$$

Then $\lim_{y\to b} g(y) = \lim_{x\to a} h(x) = L$.

Proof: Let $\epsilon > 0$. Then there exist a $\delta > 0$ so that

$$\begin{aligned} 0 < |x-a| < \delta \quad &\Rightarrow \quad |f(x,y) - g(y)| < \frac{\epsilon}{2} \\ 0 < |y-b| < \delta \quad &\Rightarrow \quad |f(x,y) - h(x)| < \frac{\epsilon}{2} \\ 0 < ||(x,y)-(a,b)|| < \delta \quad &\Rightarrow \quad |f(x,y) - L| < \frac{\epsilon}{2}. \end{aligned}$$

Proof continued...

Note that

$$|g(y) - L| \leq |g(y) - f(x,y)| + |f(x,y) - L|, \quad \forall\ x, y. \tag{1}$$

In particular, if we fix an x such that $0 < |x - a| < \frac{\delta}{\sqrt{2}}$, then for $d = \frac{\delta}{\sqrt{2}} > 0$ and for any y with $0 < |y - b| < d$, we obtain that

$$0 < ||(x,y) - (a,b)|| = \sqrt{|x-a|^2 + |y-b|^2} < \sqrt{\frac{\delta^2}{2} + \frac{\delta^2}{2}} = \delta$$

and hence

$$|f(x,y) - L| < \frac{\epsilon}{2}.$$

For the x above, $|g(y) - f(x,y)| < \frac{\epsilon}{2}$ as $0 < |x - a| < \frac{\delta}{\sqrt{2}} < \delta$. Thus, it follows from (1) that

$$|g(y) - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

whenever $0 < |y - b| < d = \frac{\delta}{\sqrt{2}}$. Therefore $\lim_{y \to b} g(y) = L$. Similarly we can show that $\lim_{x \to a} h(x) = L$. $\square$

Iterated limits continued...

Exercise 5.5: Let

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

we have

$$\lim_{x\to 0}\lim_{y\to 0}\frac{x^2y^2}{x^2y^2+(x-y)^2} = \lim_{x\to 0} 0 = 0$$

Similarly, one has $\lim_{y\to 0}\lim_{x\to 0} f(x,y) = 0$.

However, choosing $z_n = (\frac{1}{n}, \frac{1}{n})$, shows that $f(z_n) = 1$ for all $n \in \mathbb{N}$. Now choose $z_n = (\frac{1}{n}, \frac{1}{2n})$ to see that the limit cannot exist.

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 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

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

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".

In the section on Iterated Limits, we studied the following function from Exercise 5.5:

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

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 5.3(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. That is,

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

where $o(h)$ 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 curve*. 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 "$o(h)$" 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

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

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

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

where $o(h, k)$ 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 1×2 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 1×2 matrix can be multiplied by a column vector (which is 2×1 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 to be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df((x_0, y_0))$ with the property that

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

for some function $o(h, k)$ 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 1×2 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 λ. 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).

A condition for differentiability

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

Theorem 26: Let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x_0, y_0)$ 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.

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.

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 i and j the gradient can be written as

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

As one sees easily, the gradient is related to the directional derivative in the direction v.

$$\nabla_v f = \nabla f \cdot v.$$

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 Theorem 26 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 27: 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

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

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

$$f(x(t+h), y(t+h)) = f(x(t)+h[x'(t)+o_1(h)], y(t)+h[y'(t)+o_2(h)])$$

for functions o_1 and o_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

$$f(x(t), y(t)) + Df(h[x'(t) + o_1(h)], h[y'(t) + o_2(h)]) + o_3(h)h$$

by using the differentiability of f, for some other function $o_3(h)$ which goes to zero as h goes to zero (you may need to think about this step a little). In turn, this can be rewritten

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

We can now divide by h to get the desired result.

An application to tangents of curves

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

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

Question: In what direction a function f changes 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 θ 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 ∇f. In other words the function is increasing fastest in the direction v given by ∇f. Thus the unit vector v for which $\nabla_v f(x_0, y_0, z_0)$ maximizes is

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

Similarly the unit vector v for which $\nabla_v f(x_0, y_0, z_0)$ minimizes is

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

Example:

Let $f(x,y) := 4 - x^2 - y^2$ for $(x,y) \in \mathbb{R}^2$. Then f is differentiable on $\mathbb{R}^2$, and $(\nabla f)(x_0, y_0) = (-2x_0, -2y_0)$ for all $(x_0, y_0) \in \mathbb{R}^2$.

Let $(x_0, y_0) := (1,1)$. Then $(\nabla f)(1,1) = (-2,-2)$.

Consider the surface $z = f(x,y)$. At $(1,1)$,

- the direction of steepest ascent is $(-2,-2)/\|(-2,-2)\| = (-1/\sqrt{2}, -1/\sqrt{2})$,
- the direction of steepest descent is $-(-2,-2)/\|(-2,-2)\| = (1/\sqrt{2}, 1/\sqrt{2})$,

Exercise: Find the directions in which the directional derivative of $f(x, y) = x^2 + \sin xy$ at the point $(1, 0)$ has the value 1.

Solution: We compute ∇f first:

$$\nabla f(x, y) = (2x + y \cos xy, x \cos xy),$$

so at $(1, 0)$ we get, $\nabla f(1, 0) = (2, 1)$.

To find the directional derivative in the direction $v = (v_1, v_2)$ (where v is a unit vector), we simply take the dot product with the gradient:

$$\nabla_v f(1, 0) = 2v_1 + v_2.$$

This will have value "1" when $2v_1 + v_2 = 1$, subject to $v_1^2 + v_2^2 = 1$, which yields $v_1 = 0, v_2 = 1$ or $v_1 = 4/5, v_2 = -3/5$.

Gradient condition

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 $x^2 + y^2 - z^2 = 0$.

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.

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.

Conclusion: The gradient of f is perpendicular to the surface $f(x, y, z) = b$.

Review of the gradient

Exercise 3: Find $D_u F(2,2,1)$ where D_u denotes the directional derivative of the function $F(x,y,z) = 3x - 5y + 2z$ and u is the unit vector in the outward normal to the sphere $x^2 + y^2 + z^2 = 9$ at the point $(2,2,1)$.

Solution: The unit outward normal to the sphere $g(x,y,z) = 9$ at $(2,2,1)$ is given by

$$\frac{\nabla g(2,2,1)}{\|\nabla g(2,2,1)\|}.$$

We see that $\nabla g(2,1,1) = (4,4,2)$ so the corresponding unit vector is $(2,2,1)/3$.

To get the directional derivative we simply take the dot product of ∇F with u:

$$(3,-5,2) \cdot (2,2,1)/3 = -\frac{2}{3}.$$

Review - problems involving the gradient

Exercise Find the points on the hyperboloid $x^2 - y^2 + 2z^2 = 1$ where the normal line is parallel to the line that joins the points $(3, -1, 0)$ and $(5, 3, 6)$.

Solution: The hyperboloid is an implicitly definined surface. A normal vector at a point (x_0, y_0, z_0) on the hyperboloid is given by the gradient of the function $x^2 - y^2 + 2z^2$ at (x_0, y_0, z_0):

$$\nabla f(x_0, y_0, z_0) = (2x_0, -2y_0, 4z_0).$$

We require this vector to be parallel to the line joining the points $(3, -1, 0)$ and $(5, 3, 6)$. This line lies in the same direction as the vector $(5 - 3, 3 + 1, 6 - 0) = (2, 4, 6)$. Thus we need only solve the equations

$$2x_0 = 1, \quad -2y_0 = 4, \quad 4z_0 = 6,$$

which give $x_0 = 1/2$, $y_0 = -2$ and $z_0 = 3/2$. Thus, we need to find λ such that $\lambda(1/2, -2, 3/3)$ lies on the hyperboloid. Substituting in the equation yields $\lambda = \pm\sqrt{2/3}$.

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

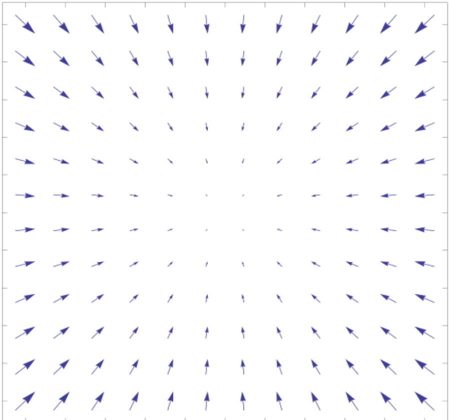
Vector fields

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 $m = n = 3$.

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?.

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.



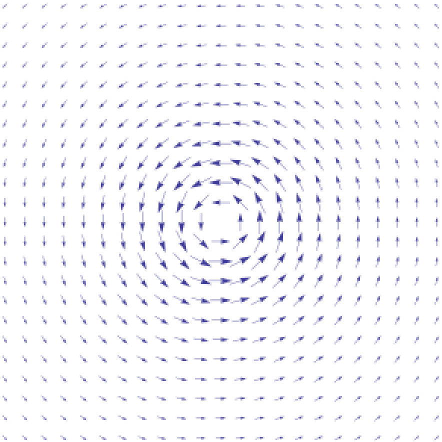
What function from $\mathbb{R}^2$ to $\mathbb{R}^2$ does this picture represent?

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

the radial vector field.

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

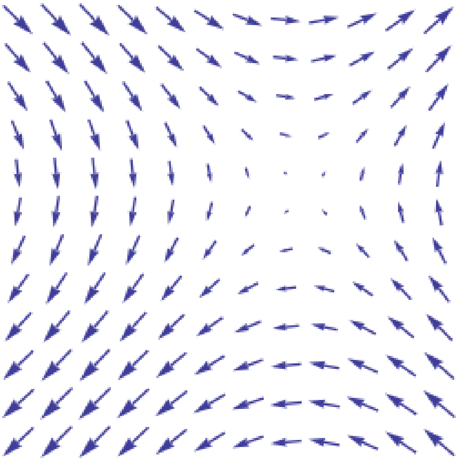
How about this one?



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

This is an example of an irrotational vector field. It cannot be written as the gradient of a potential function.

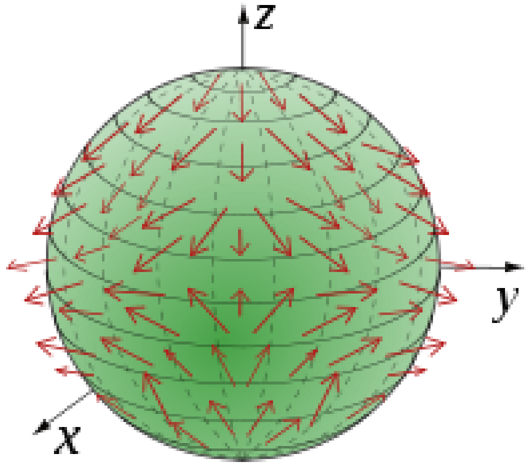
Here is another (more complicated one)



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

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

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.

The derivative for $f : U \to \mathbb{R}^n$

We now define the derivative for a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$.

The function f is said to be differentiable at a point x if there exists a an $n \times m$ matrix $Df(x)$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - Df(x) \cdot h\|}{\|h\|} = 0.$$

Here $x = (x_1, x_2, \ldots, x_m)$ and $h = (h_1, h_2, \ldots h_m)$ are vectors in $\mathbb{R}^m$.

The matrix $Df(x)$ is usually called the total derivative of f. It is also referred to as the Jacobian matrix. What are its entries?

From our experience in the 2×1 case we might guess (correctly!) that the entries will be the partial derivatives.

Here is the total derivative or the derivative matrix written out fully.

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_m}(x) \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & \cdots & \frac{\partial f_n}{\partial x_m}(x) \end{pmatrix}$$

In the 2×2 case we get

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{pmatrix}.$$

As before, the derivative may be viewed as a linear map, this time from $\mathbb{R}^m$ to $\mathbb{R}^n$ (or, in the case just above, from $\mathbb{R}^2$ to $\mathbb{R}^2$).

Rules for the total derivative

Just like in the one variable case, it is easy to prove that

$$D(f+g)(x) = Df(x) + Dg(x).$$

Somewhat harder, but only because the notation gets more cumbersome, is the Chain rule:

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x),$$

where $\circ$ on the right hand side denotes matrix multiplication.

Theorem 26 holds in this greater generality - a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ is differentiable at a point x_0 if all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ $1 \leq i \leq n$, $1 \leq j \leq m$, are continuous in a neighborhood of x_0 (define a neighborhood of x_0 in $\mathbb{R}^m$!).

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 28: 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. Theorem 28 says that for $\mathcal{C}^2$ functions, the order in which one takes partial derivatives does not matter.

From now on we will 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_k times with respect to x_k, then n_{k-1} times with respect to x_{k-1}, 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.

Recall that 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. Theorem 28 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_k times with respect to x_k, then n_{k-1} times with respect to x_{k-1}, and so on. The number n is nothing but $n_1 + n_2 + \dots + n_k$. 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 $f_1, f_2, \dots, f_n$ 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. We will return to this later.

Recall that we often write f_x instead of $\frac{\partial f}{\partial x}$ for the partial derivative of f with respect to x. Similarly, we will sometimes write f_{xx} for $\frac{\partial^2 f}{\partial x^2}$ or f_{xy} for $\frac{\partial^2 f}{\partial y \partial x}$ etc. when convenient.

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 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 29: 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 do this we will need to look at 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 as 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.

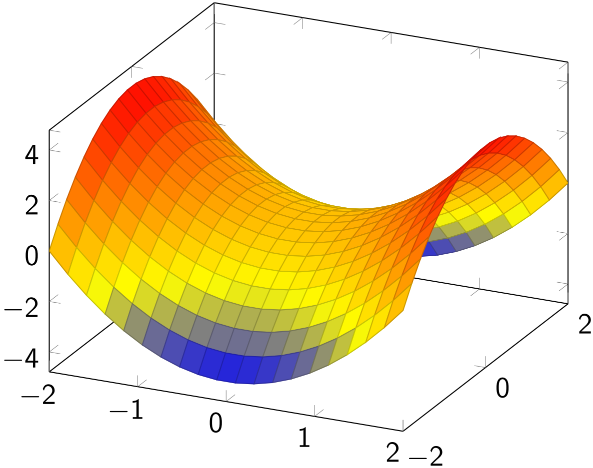
We will explain this after stating the following theorem.

Theorem 30: 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

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

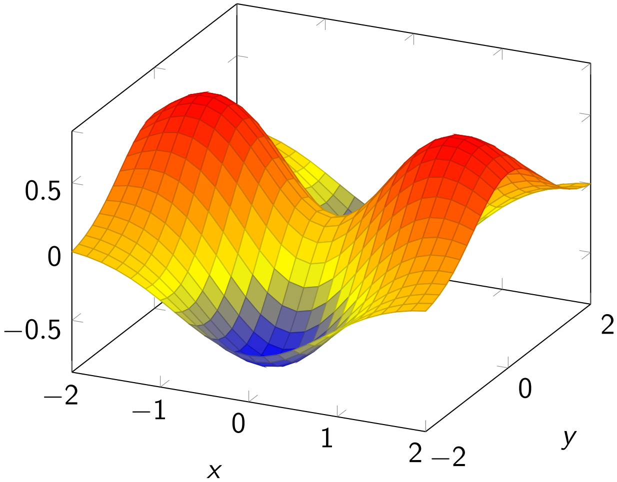
Next we have to find the partial derivatives of order 2. We have

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

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

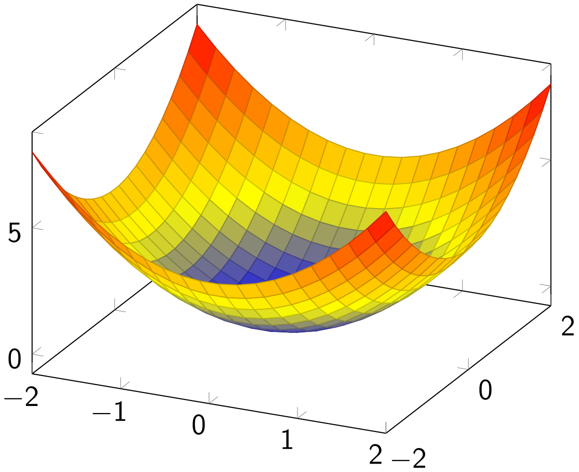
$$z = g(x, y) = Ax^2 + 2Bxy + Cy^2.$$

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:

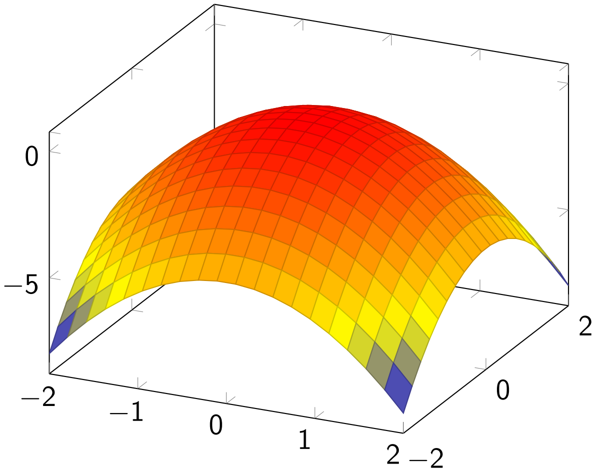
1. If $AC - B^2 > 0$, the function g has a local minimum if $A > 0$ and a local maximum if $A < 0$.
2. If $AC - B^2 < 0$, the function g has a saddle point, that is, in a small disc around the point, the function does not lie on any one side of its tangent plane.

A local minimum



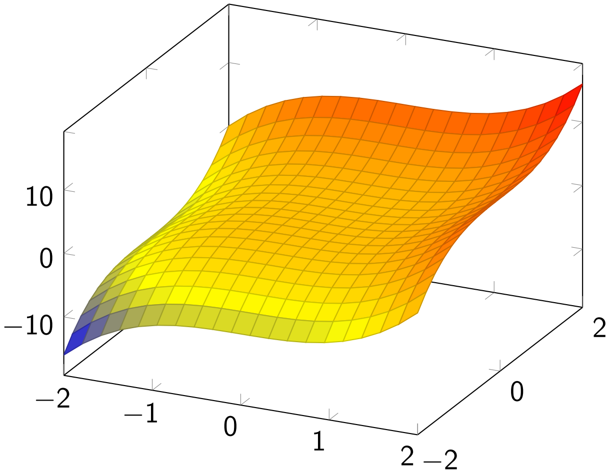
The graph of $x^2 + y^2$ has a local minimum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A > 0$.

A local maximum



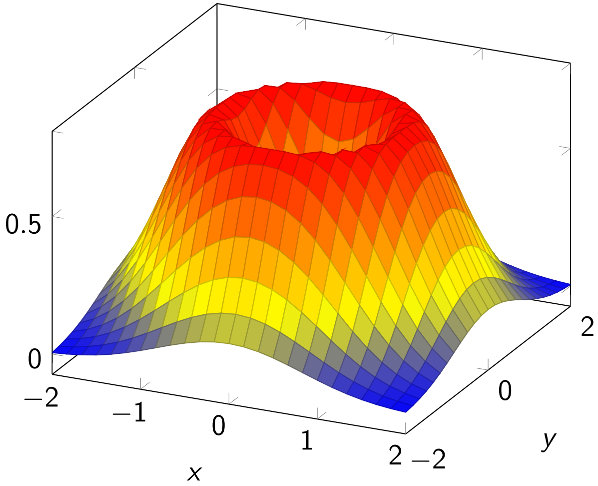
The graph of $-x^2 - y^2$ has a local maximum at $(0, 0)$. Clearly $AC - B^2 = 1 > 0$ and $A < 0$.

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.

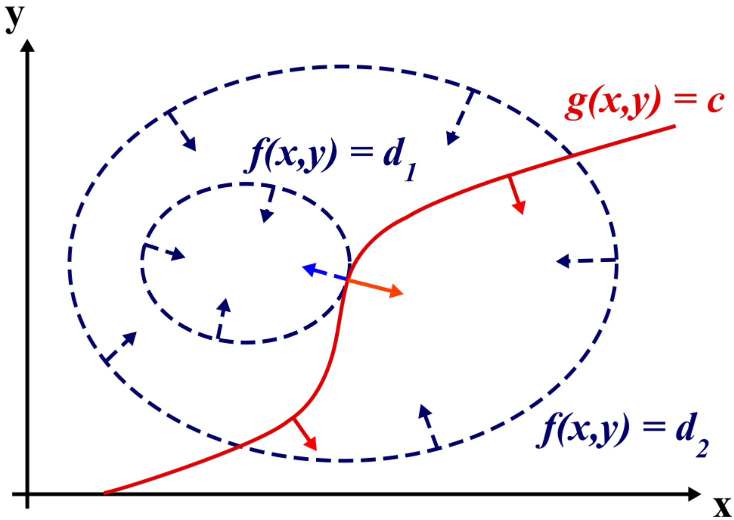
The two variable problem

Suppose we are given a function $f(x, y)$ in two variables. We would like maximize or minimize it subject to the constraint that $g(x, y) = c$. In geometric terms, we want to find the maximum or minimum values of f while stating on the curve $g(x, y) = c$.

Where might we find these maxima or minima?

One way of doing it would be to walk along the curve $g(x, y) - c$ meeting the level curves $f(x, y) = d$ as d increases. The smallest value of d (if it exists!) such that $g(x, y) = c$ and $g(x, y) = d$ meet will surely be the minimum value for the function $f(x, y)$ on the curve $g(x, y) = c$. What special property holds at this point?

In pictures



http://en.wikipedia.org/wiki/Lagrange_multiplier#mediaviewer/File:LagrangeM

The condition on the normals

From the picture one might guess that at the value d_1 the curves are tangent to each other, or, equivalently, their normals are parallel to each other. Why is this?

Suppose that this were not the case. Recall that the normals to the level curves of f are the directions in which f is decreasing on increasing the fastest. If this normal is not perpendicular to the curve $g(x, y) = c$, then it will have a component tangent to this curve.

Hence, the function $f(x, y)$ will be either increasing or decreasing at the point of intersection of the two curves as one goes along $g(x, y) = c$. This shows that the point of intersection cannot be an extreme point.

The gradient condition

From the discussion in the previous slide, it follows that the normals of the curves $f(x,y) = d$ and $g(x,y) = c$ must be parallel to each other at the extrema. We are thus looking for points (x_0, y_0) such that

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0),$$

subject to the constraint condition.

$$g(x,y) = c.$$

The λ that occurs above is called the Lagrange multiplier. The points (x_0, y_0) satisfying the above conditions are critical points. We have to actually check whether they are extrema or not.

Checking for the extrema

To check for extrema all we have to do is to evaluate the function $f(x, y)$ at each of points (x_0, y_0) that we find after solving the equation above. This will enable us to decide whether a given point is a maximum or minimum.

Exercise: Find the maximum and minimum values of the function $f(x, y) = x^2 + 2y^2$ that lie on the circle $x^2 + y^2 = 1$.

Solution: We know that we must solve the equations

$$\nabla f = \lambda \nabla g \quad \text{and} \quad x^2 + y^2 = 1,$$

where $g(x, y) = x^2 + y^2$. The first equation yields the pair of equations

$$2x = \lambda \cdot 2x \quad \text{and} \quad 4y = \lambda \cdot 2y.$$

The first equation above yields $x = 0$ or $\lambda = 1$. If $x = 0$, we must have $y = \pm 1$. If $\lambda = 1$, then $4y = 2y$ in the second equation, so $y = 0$ and $x = \pm 1$.

Thus we must evaluate $f(x,y)$ at the points $(\pm 1, 0)$ and $(0, \pm 1)$.
We obtain $f(1,0) = 1,\ f(-1,0) = 1,\ f(0,1) = 2,\ f(0,-1) = 2.$
Thus the maximum value is 2 and the minimum 1.

The three variable problem

The same reasoning as before applies to the three variable constrained problem, that is, to find the maxima and minima of a function $f(x, y, z)$ subject to $g(x, y, z) = c$. If we are moving on the surface of g, this means that at any point, we are in the tangent plane of g. On the other hand, the directions in which f is stationary, are those perpendicular to its gradient.

It follows that we must once again solve the equations

$$\nabla f = \lambda \nabla g \quad \text{and} \quad g(x, y, z) = c,$$

to obtain the extreme points, and then evaluate the function $f(x, y, z)$ at these points.

Exercise: Find the dimensions of the box with the largest volume having a surface area of $150 cm^2$.

Solution: The volume of the box is given by $f(x, y, z) = xyz$. This has to be maximized subject to the constraint that $g(x, y, z) = 2xy + 2yz + 2zx = 150$.

The condition $\nabla f = \lambda \nabla g$ yields

$$yz = \lambda(2y + 2z), \; zx = \lambda(2z + 2x), \; xy = \lambda(2x + 2y).$$

while the constraint condition give $xy + yx + zx = 75$.
Multiplying the previous equations by x, y and z respectively we get

$$xyz = 2\lambda x(y + z), \; xyz = 2\lambda y(z + x), \; xyz = 2\lambda z(x + y).$$

Equating the first two gives $\lambda = 0$ or $x = y$. But if $\lambda = 0$ we get $y = 0$ or $z = 0$, which is not possible. The equations above are symmetric in x, y and z, so if $x = y$ we must have $x = y = z$. This yields $3x^2 = 75$ or $x = y = z = 5$.

Taylor's theorem in two variables

If we look at the quadratic surface $z = f(x, y) = Ax^2 + 2Bxy + Cy^2$, 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 (recall that $D = 4AC - 4B^2 = 4(AC - b^2))$.

Why is this true?

The answer lies in a two variable form of Taylor's Theorem:

Theorem 32: 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 quadratic 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 (as $z_0 = f(x_0, y_0)$ is fixed), 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, where $h = b - a$. 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.

Solution (Proof of Taylor's theorem in two variable):
For $g(t) = f(tx + (1 - t)x_0, ty + (1 - t)y_0)$, we obtain (by the chain rule) that

$$g'(t) = \frac{\partial f}{\partial x}(tx + (1 - t)x_0, ty + (1 - t)y_0)(x - x_0)$$

$$+\frac{\partial f}{\partial y}(tx + (1 - t)x_0, ty + (1 - t)y_0)(y - y_0)$$

$$\implies g'(0) = \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

and

$$
\begin{aligned}
g''(t) = {} & \frac{\partial^2 f}{\partial x^2}(tx + (1-t)x_0, ty + (1-t)y_0)(x - x_0)^2 \\
& + \frac{\partial^2 f}{\partial y \partial x}(tx + (1-t)x_0, ty + (1-t)y_0)(x - x_0)(y - y_0) \\
& + \frac{\partial^2 f}{\partial y^2}(tx + (1-t)x_0, ty + (1-t)y_0)(y - y_0)^2 \\
& + \frac{\partial^2 f}{\partial x \partial y}(tx + (1-t)x_0, ty + (1-t)y_0)(x - x_0)(y - y_0)(x - x_0)
\end{aligned}
$$

$$
\begin{aligned}
\implies g''(0) = {} & \frac{\partial^2 f}{\partial x^2}(x_0, y_0)(x - x_0)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)(x - x_0)(y - y_0) \\
& + \frac{\partial^2 f}{\partial y^2}(x_0, y_0)(y - y_0)^2.
\end{aligned}
$$

Thus, for $(x, y) = (x_0 + h, y_0 + k)$,

$$g'(0) = f_x(x_0, y_0)h + f_y(x_0, y_0)k$$

and

$$g''(0) = f_{xx}(x_0, y_0)h^2 + 2f_{xy}(x_0, y_0)hk + f_{yy}(x_0, y_0)k^2.$$

Now by the Taylor's theorem in one variable,

$$g(1) = g(0) + g'(0)1 + \frac{g^{(2)}(0)}{2!}1^2 + \frac{1}{2!}[g^{(2)}(c) - g^{(2)}(0)]1^2$$

for some $c \in (0, 1)$. Thus, we obtain that

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + [f_x(x_0, y_0)h + f_y(x_0, y_0)k]$$

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

where

$$\tilde{R}_2(h, k) = \frac{1}{2!}[g^{(2)}(c) - g^{(2)}(0)]$$

$$= \frac{1}{2!}[(f_{xx}(x_0+ch, y_0+ck) - f_{xx}(x_0, y_0))h^2 + 2(f_{xy}(x_0+ch, y_0+ck) - f_{xy}(x_0, y_0))hk$$

$$+(f_{yy}(x_0 + ch, y_0 + ck) - f_{yy}(x_0, y_0))k^2]$$

and hence

$$\lim_{(h,k)\to(0,0)} \frac{\tilde{R}_2(h, k)}{(||(h, k)||)^2} = 0.$$

$\square$

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

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

(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 $x^2 - y^2 > 16$, $z \leq 1/2$, 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 $\mathbb{R}^2$.

Rectangles

Any rectangle R in the plane can be described as the set of points in the cartesian product $[a, b] \times [c, d]$ of two closed intervals.

We can define the notion of a partition of a rectangle. How can this be done?

The easiest way is to take a partition P_1 of $[a, b]$ and a partition P_2 of $[c, d]$ and take the product of the two partitions. Thus if

$$P_1 = \{a = x_0, x_1 \ldots, x_m = b)\} \quad \text{and} \quad P_2 = \{c = y_0, y_1, \ldots, y_n = d\},$$

we take the collection of points

$$P = \{(x_i, y_j) \,|\, 1 \leq i \leq m,\, 1 \leq j \leq n\}.$$

The point (x_i, y_j) is the left bottom corner of the rectangle $R_{ij} = (x_i, x_{i+1}) \times (y_j, y_{j+1})$. As i and j vary, we get a family of rectangles R_{ij}, $0 \leq i \leq m-1$, $0 \leq j \leq n-1$. By identifying each rectangle with its left bottom corner we can think of P as the collection of these rectangles R_{ij}. Clearly, $R = \cup_{i,j} R_{ij}$, and the collection of rectangles P is called a partition of R.

Partitions of rectangles and the integral

You may recall that we have several equivalent definitions of Riemann integration in one variable. We can model the notion of Riemann integration in two variables on any one of those. Clearly, the notion of upper and lower sums associated to a partition makes sense so we could define the Darboux integral as before.

We set $\Delta_{ij} = (x_{i+1} - x_i) \times (y_{j+1} - y_j)$, the area of the rectangle R_{ij}.

If $f : R \to \mathbb{R}$ is a function of two variables, define

$$m_{ij} = \inf_{(x,y) \in R_{ij}} f(x, y)$$

and the lower sum of f associated to the partition P by.

$$L(f, P) = \sum_{i,j}^{m-1,n-1} m_{ij} \Delta_{ij}.$$

We can similarly define the upper sums, and now it is clear that the lower and upper integrals, and hence also the Darboux integral can be defined.

The Riemann integral

A second approach arises by defining the norm of a partition of a rectangle. Again, this is easy. In the one variable case the norm of a partition was simply the length of the largest of the sub-intervals. What is the analogue for rectangles? Clearly, we could define

$$\|P\|_1 = \max_{0 \le i \le m-1, 0 \le j \le n-1} (x_{i+1} - x_i) \times (y_{j+1} - y_j).$$

As before, we can define a tagged partition (P, t), where $t = \{t_{ij}\}_{i,j}$ is a collection of points such that $t_{ij} \in R_{ij}$.

The Riemann sum S associated to (P, t) is defined by

$$S(f, P, t) = \sum_{i,j}^{m-1,n-1} f(t_{ij}) \Delta_{ij}.$$

Now it should be clear how to define the Riemann integral: it is a number S such that for any $\epsilon > 0$, there is a δ such that

$$|S(f, P, t) - S| < \epsilon,$$

for every tagged partition (P, t) with the property that $\|P\|_1 \le \delta$.

One problem with the previous definition

There is one problem with the definition made in the last slide. The norm that we have selected for partitions is not such a good one.

For instance, one could take $P_1 = \{0 < 1/2 < 1\}$ and $P_{2,n} = \{0 < 1/n, 2/n, \ldots < 1\}$ and take $P_n = P_1 \times P_{2,n}$ as a partition of the unit square. In this case, clearly $\|P\| = 1/2n$ goes to zero as $n \to \infty$.

Exercise 1: Find a function $f(x, y)$ on the unit square and an $\epsilon > 0$ such that

$$|S(f, P_n, t) - V| > \epsilon,$$

for any tag t of P_n, where V is the volume of the solid region lying above the unit square and below the graph of $z = f(x, y)$ (note, there are very simple functions with the property).

Remedy

How to fix the problem? The problem is that the rectangles in our partition may be very thin in one direction but remain fat in the other one.

In order to avoid such situations, we have to change the definition of the norm that we have given. We have to select a norm so that when the norm of P is small, both sides of our rectangles are guaranteed to be small.

We define $\|P\| = \max_{i,j}\{(x_{i+1} - x_i), (y_{j+1} - y_j)\}$. Clearly this norm has the desired property and is the correct analogue of the norm of a partition in one variable.

Thus, to get the correct definition of a Riemann integral for a two variable function one must replace $\|P\|_1$ in the definition given above by $\|P\|$.

Regular Partitions

We will not use either of the above approaches, preferring instead the third approach. Recall that in Definition 2 of the one variable integral we saw that it was enough to restrict our attention to a fixed family of partitions. This is what we will do, taking a particularly simple family of partitions.

The regular partition of R of order n is is partition defined inductively by $x_0 = a$ and $y_0 = c$ and

$$x_{i+1} = x_i + \frac{b-a}{n} \quad \text{and} \quad y_{j+1} = y_j + \frac{d-c}{n},$$

$1 \leq i, j \leq n-1$. We take $t = \{t_{ij} \in R_{ij}\}$ to be an arbitrary tag.

Definition: We say that the function $f : R \to \mathbb{R}$ is Riemann integrable if the Riemann sum

$$S(f, P_n, t) = \sum_{i,j=0}^{n-1} f(t_{ij})\Delta_{ij}$$

tends to a limit S for any choice of tag t.

The Riemann integral continued

This limit value is usually denoted as

$$\int\int_R f, \quad \int\int_R f(x,y)dA, \quad \text{or} \quad \int\int_R f(x,y)dxdy.$$

The preceding definition is sometimes rewritten as

$$\lim_{n\to\infty} \sum_{i,j=0}^{n-1} f(t_{ij})\Delta_{ij} = \int\int_R f.$$

If $f(x,y) \geq 0$ for all values of x and y, then the Riemann integral has a geometric interpretation. It is obviously the volume of the region under the graph of the function $z = f(x,y)$ and above the rectangle R in xy-plane.

The integral may also be interpreted as mass in some physical situations; for example, if we have a rectangular plate and $f(x,y)$ represents the density of the plate at a given point, then the integral above gives the mass of the whole plate.

The main theorem

In the one variable case, we saw that a bounded function with at most a finite number of discontinuities on a closed bounded interval is Riemann integrable. The reason that a finite number of discontinuities do not matter is that points have length zero. What might be the analogous result in two variables?

In two variables the geometry of the set of points of discontinuity can be more complicated. Still, what are the analogues of points in this case? In other words what sets have "zero area"?

Theorem 33: If a function f is bounded and continuous on R except possibly along a finite number of graphs of $\mathcal{C}^1$ functions, then f is integrable on R.