

Chapter 2. Functions of several variables

This material is covered in Thomas (chapter 14 in the 11th edition, or chapter 11 in the 10th edition). On the web site <http://www.maths.tcd.ie/~richardt/2E1> there will be a more detailed explanation of the sections of those chapters we are covering. We start with planes (and more on lines), which are actually earlier in the book (section 12.5 in the 11th edition, 10.3 in the 10th) as we need to be comfortable manipulating these. It also helps develop an intuition for working with concepts in space (\mathbb{R}^3).

2.1 Planes in space. The basic fact is that a plane in \mathbb{R}^3 has an equation

$$\alpha x + \beta y + \gamma z = c$$

where (α, β, γ) is a (fixed) nonzero vector in \mathbb{R}^3 perpendicular to the plane and c is a constant.

We can think of equations of the above type, linear equations in 3 variables, as the simplest kind of equation. Recall that in 2 variables lines are the most basic shapes, and graphs of linear functions $y = mx + c$ are lines. We are used to thinking of m as the slope and c as the y -intercept (so that $(0, c)$ is a point on the line). The slope m gives us the direction of the line, and to distinguish one specific line among all (parallel) lines with the same slope we need to know one point on the line.

When we look at 3 variables, it is a little simpler to see that we get a plane as the solutions of a linear equation when we take the case where $c = 0$, or look at a homogeneous equation

$$\alpha x + \beta y + \gamma z = 0.$$

We can recognise the left hand side as the dot product $(\alpha, \beta, \gamma) \cdot (x, y, z)$ and then we can interpret the equation that the dot product is 0 in terms of perpendicular vectors: $(\alpha, \beta, \gamma) \perp (x, y, z)$.

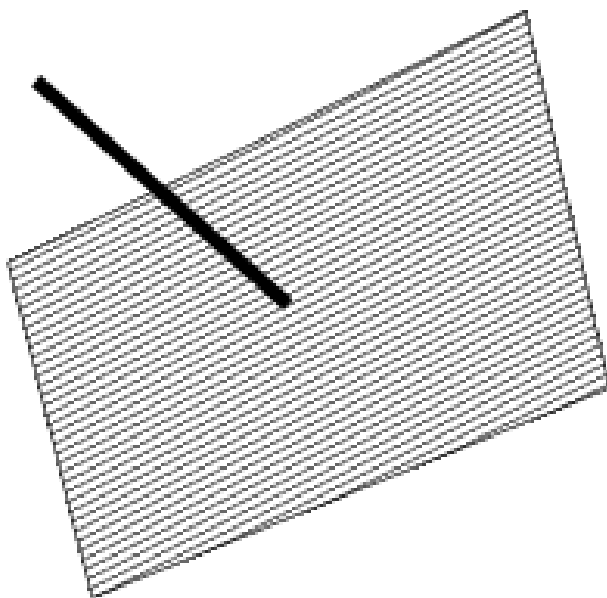
If we now think geometrically, we see that the points (x, y, z) whose position vectors are perpendicular to the fixed vector (α, β, γ) are those points which lie in the plane through the origin $(0, 0, 0)$ perpendicular to (α, β, γ) . So the equation $\alpha x + \beta y + \gamma z = 0$ does describe a plane through the origin.

On the other hand, starting with any plane through the origin, we can take a nonzero vector (α, β, γ) perpendicular to the plane and the equations $\alpha x + \beta y + \gamma z = 0$ will describe the plane we began with.

Before explaining why we get all planes (not necessarily through the origin) by allowing c to be nonzero, we consider the question of how we might describe the orientation of a plane in space. For lines in the plane we use the slope usually (though there are vertical lines $x = a$ that have no slope). We could perhaps think about the slope of a plane as the slope of the steepest line up the plane, but this will not tell us the orientation of our plane in space. We can move the plane around (say rotating it around a vertical direction), keeping the same largest slope, but changing the way the plane is oriented in space.

If you think about it for a moment, you will see that the method of picking a vector perpendicular to the plane (which we sometimes call a *normal vector* to the plane) works well. You could maybe construct a little model, like plasterers use, or a flat board with a stick handle perpendicular to the board (attached by glue or a screw to one side of the board). If you want to

move the board around parallel to itself you have to keep the stick handle pointing in a constant direction. (To be more precise, you could also turn the handle around the other way, so that it will reverse direction completely, and still have the board in the same plane.)

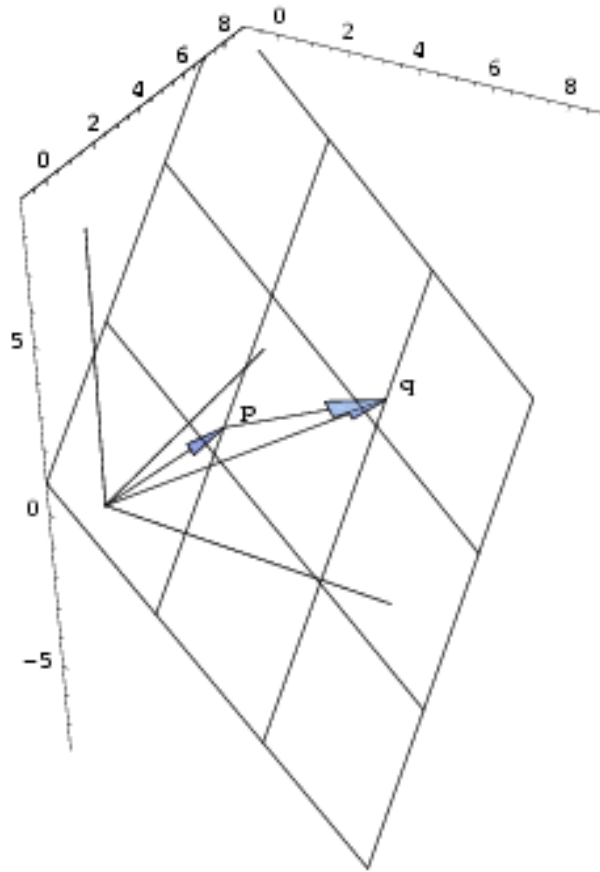


You should see that picking a vector perpendicular to it allows you to say which way a plane is tilted in space. The normal vector is not unique — multiplying the vector by a positive or negative (but not zero) scalar will not change the orientation of a perpendicular plane.

Now if we think about all the planes perpendicular to a given (or fixed) normal vector, then we can see that we could pick out one of these many planes by knowing one point on the plane.

Say \mathbf{p} is a point on our plane, \mathbf{n} is a normal vector. How do we come up with an equation that is satisfied by all the points on the plane? And of course we want those points that satisfy the equation to be on our plane. We have the solution in case the given point $\mathbf{p} = \mathbf{0}$ is the origin. If \mathbf{q} is on the plane, then $\mathbf{n} \cdot \mathbf{q} = 0$.

In general, the position vectors of points \mathbf{q} on the plane through \mathbf{p} perpendicular to \mathbf{n} would not have to be perpendicular to \mathbf{n} . However, the difference vector $\mathbf{q} - \mathbf{p}$ is represented by the arrow (lying in the plane) from \mathbf{p} to \mathbf{q} , and so must be perpendicular to \mathbf{n} . This is be a sketch of a plane, the 3 axes and an example of the \mathbf{p} and \mathbf{q} .



It should be clear that if any point $\mathbf{q} \in \mathbb{R}^3$ satisfies $\mathbf{q} - \mathbf{p}$ perpendicular to \mathbf{n} , then \mathbf{q} is in the plane. So the plane has equation

$$\mathbf{n} \cdot (\mathbf{q} - \mathbf{p}) = 0,$$

or

$$\mathbf{n} \cdot \mathbf{q} = \mathbf{n} \cdot \mathbf{p}.$$

If we write $c = \mathbf{n} \cdot \mathbf{p}$ (which is a constant depending on the chosen point \mathbf{p} and the chosen normal \mathbf{n}), then the equation has the form

$$\mathbf{n} \cdot \mathbf{q} = c.$$

Writing $\mathbf{n} = (\alpha, \beta, \gamma)$ and $\mathbf{q} = (x, y, z)$ for the arbitrary point on the plane we get the equation

$$\alpha x + \beta y + \gamma z = c,$$

as claimed.

2.2 Example. Find parametric equations for the line of intersection of the two planes

$$\begin{aligned} 2x - 3y + 4z &= 0 \\ -x + y + 2z &= 3 \end{aligned}$$

We can write the parametric equations of the line if we know one point \mathbf{a} on the line and a (nonzero) vector \mathbf{b} parallel to the line.

The normal vectors to the planes are $(2, -3, 4)$ and $(-1, 1, 2)$. Thus the line is perpendicular to both vectors, or parallel to their cross product. We can take

$$\mathbf{b} = (2, -3, 4) \times (-1, 1, 2).$$

To find a point \mathbf{a} on the line, we could (for instance) look for the point on the line where $z = 0$, which we would find by solving 3 simultaneous linear equations (or the system of 3 equations) got by taking the given equations for the two planes and the extra equation $z = 0$.

2.3 Cartesian equations of lines. If we start with any line in space \mathbb{R}^3 , there are many planes that contain the line. If you take one plane containing the line and rotate it around the line, you get all the other planes containing that line.

So if you want to represent the line as the intersection of two planes, there are many possibilities. (What this means is to kind of reverse the example above, where we started with two planes and found parametric equations for the line where they intersect.) There is however a way to pick two planes in a standard way. Starting with parametric equations for a line

$$\begin{aligned}x &= a_1 + b_1 t \\y &= a_2 + b_2 t \\z &= a_3 + b_3 t\end{aligned}$$

we can solve each of the three equations for the parameter t in terms of all the rest of the quantities. We get

$$t = \frac{x - a_1}{b_1}, \quad t = \frac{y - a_2}{b_2}, \quad t = \frac{z - a_3}{b_3}.$$

(At least we can do this if none of b_1 , b_2 or b_3 is zero.) If (x, y, z) is a point on the line, there must be one value of t that gives the point (x, y, z) from the parametric equations, and so the 3 values for t must coincide. We get

$$\frac{x - a_1}{b_1} = \frac{y - a_2}{b_2} = \frac{z - a_3}{b_3}.$$

These equations are called *cartesian equations for the line*.

You might wonder what sort of equation we have with two equalities in it? Well it is two equations

$$\frac{x - a_1}{b_1} = \frac{y - a_2}{b_2} \text{ and } \frac{y - a_2}{b_2} = \frac{z - a_3}{b_3}.$$

Each of these two equations is a linear equation, and so the equation of a plane. So we are representing the line as the intersection of two particular planes when we write the cartesian equations.

The first plane

$$\frac{x - a_1}{b_1} = \frac{y - a_2}{b_2}$$

could be rewritten

$$\frac{1}{b_1}x - \frac{1}{b_2}y + 0z = \frac{a_1}{b_1} - \frac{a_2}{b_2}$$

so that it looks like $\alpha x + \beta y + \gamma z = c$. We see that there is no z term in the equation, or the normal vector $(1/b_1, -1/b_2, 0)$ is horizontal. This means that the plane is parallel to the z -axis (or is the vertical plane that contains the line we started with).

We could similarly figure out that the second plane

$$\frac{y - a_2}{b_2} = \frac{z - a_3}{b_3}$$

is the plane parallel to the x -axis that contains our line.

Now, there is an aspect of this that is a bit untidy. From the cartesian equations we can see that

$$\frac{x - a_1}{b_1} = \frac{z - a_3}{b_3}$$

follows immediately, and this is the equation of the plane parallel to the y -axis that contains our line. It seems unsatisfactory that we should pick out the two planes parallel to the z -axis and the x -axis that contains our line, while discriminating against the y -axis in a rather arbitrary way. So although the cartesian equations reduce to representing our line as the intersection of two planes, we could take the two planes to be any two of the planes containing the line and parallel to the x , y and z -axes.

2.4 Example. Suppose we are given two vectors $\mathbf{A} = (3, 4, 5)$ and $\mathbf{B} = (1, -1, -2)$ that are not parallel. Then when we take linear combinations

$$\mathbf{x} = t\mathbf{A} + s\mathbf{B} \quad (\text{for all } t, s \in \mathbb{R})$$

we end up with all the points in a plane through the origin, the plane through the origin parallel to both vectors \mathbf{A} and \mathbf{B} . (To see this, work graphically. $t\mathbf{A}$ is \mathbf{A} scaled by an amount t , $s\mathbf{B}$ is \mathbf{B} scaled by s , and then we find $t\mathbf{A} + s\mathbf{B}$ by completing a parallelogram in the plane of \mathbf{A} and \mathbf{B} , and then taking the diagonal of the parallelogram from the origin.)

Let us ask for the equation of that plane.

We need a vector perpendicular to the plane and one such vector is the cross product

$$\begin{aligned} \mathbf{A} \times \mathbf{B} &= (3, 4, 5) \times (1, -1, 2) \\ &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & 4 & 5 \\ 1 & -1 & 2 \end{pmatrix} \\ &= \mathbf{i} \det \begin{pmatrix} 4 & 5 \\ -1 & 2 \end{pmatrix} - \mathbf{j} \det \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} + \mathbf{k} \det \begin{pmatrix} 2 & 4 \\ 1 & -1 \end{pmatrix} \\ &= 13\mathbf{i} - \mathbf{j} - 6\mathbf{k} \\ &= (13, -1, -6) \end{aligned}$$

Since $(0, 0, 0)$ is a point on the plane, the equation must be

$$13x - y - 6z = 0.$$

2.5 Remark. It may be instructive to relate the way we deal with planes in \mathbb{R}^3 to what would happen if we dealt with lines in \mathbb{R}^2 in the ‘same’ way.

We know that we usually write the equation of a line in the plane \mathbb{R}^2 as $y = mx + c$. However, there are vertical lines $x = a$ that have no slope and can’t be written in the form $y = mx + c$. What we could do, if we wanted to deal with lines that might be vertical, is to rewrite $y = mx + c$ in the form $-mx + y = c$. This way we can write the equation of every line in the form

$$\alpha x + \beta y = c$$

where $(\alpha, \beta) = (-m, 1)$ for lines that have a slope m , and $(\alpha, \beta) = (1, 0)$ for vertical lines. We can see then that the vector (α, β) is always perpendicular to the line. [The slope of $(\alpha, \beta) = (-m, 1)$ is $1/(-m)$ and the product of the slope m and $-1/m$ is -1 , showing that the vector is perpendicular. In the case of a vertical line, you can see that $(1, 0)$ is a horizontal vector.]

The solutions of a single linear equation $\alpha x + \beta y = c$ in \mathbb{R}^2 form a line, while the solutions of a single linear equation $\alpha x + \beta y + \gamma z = c$ in \mathbb{R}^3 form a plane. [We have to rule out the case where the coefficients on the left are all 0. $0 = c$ won’t be a line or plane.]

2.6 Functions of two variables. We now proceed to the main topic of this chapter, functions of 2 or more variables. We will deal with scalar valued functions of two variables first.

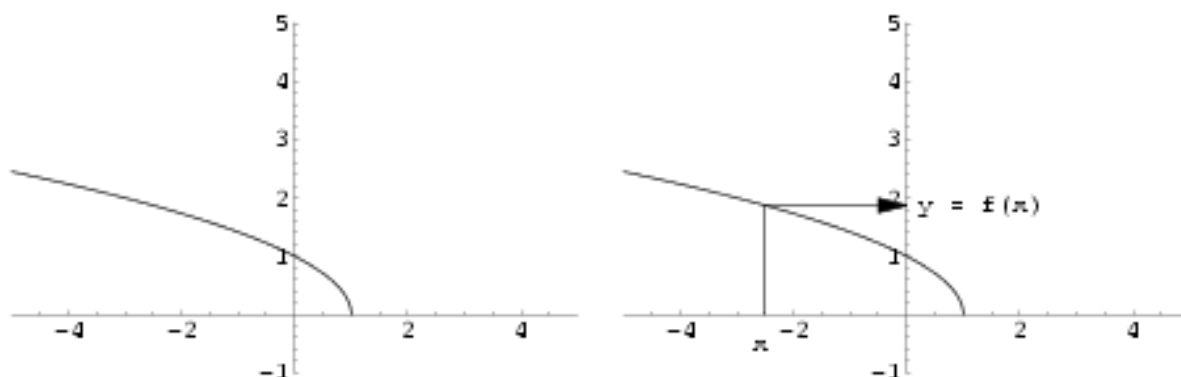
In the previous chapter we dealt with vector valued functions (or \mathbb{R}^n -valued functions) of a single variable and we could picture these as curves in \mathbb{R}^n .

You might recall that a function $f: A \rightarrow B$ from a set A (called the *domain* of the function) to a set B is a rule that assigns one and only one value in B to each element $a \in A$. So there is one value $f(a)$, which is an element of B , for each $a \in A$. The name for the set B , where the values are found, is less used. We might call it the ‘target set’ (or the fancier ‘co-domain’), as the word ‘range’ is used for the set of actual values of f .

In 1E1 you learned about calculus of functions of one real variable, scalar valued functions $y = f(x)$ where the domain was either all of \mathbb{R} or a subset of \mathbb{R} . [Often the domain was an interval or the set \mathbb{R} with one bad point removed.] We are so used to seeing examples of functions given by a formula (like $f(x) = 1/x$ or $f(x) = x^2$ or $f(x) = x \cos x + x^2 + e^x$) that we can forget that there does not have to be a simple formula to say what $f(x)$ is. There does have to be a specification of what $f(x)$ is for each x in the domain and that specification cannot be ambiguous. Only one value $f(x)$ for each x in the domain.

Typically we picture functions of one variable by looking at their graph, and then we see a curve $y = f(x)$ in \mathbb{R}^2 . The graph of a function has one special property — no vertical line can cross it more than once. The graph $y = \sqrt{1-x}$ (by the square root we **always** mean the positive square root) is an example. We can infer from the fact that we are not dealing with complex numbers that we must have the domain $A = \{x \in \mathbb{R} : x \leq 1\} = (-\infty, 1]$, or perhaps we could have some reason to restrict the domain to a smaller set than the numbers ≤ 1 .

In the following graphs, we see the graph $y = \sqrt{1-x}$ and an annotated version showing that points x in the domain, drawn along the horizontal axis, have values $y = f(x) = \sqrt{1-x}$ represented by the height of the point on the graph above x , or the point y on the vertical axis. You can think of the rule behind the function as associating y to x , or sending x to y , and the arrow represents this.



By a (scalar-valued) function of two variables, we mean a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, or a function $f: A \rightarrow \mathbb{R}$ where $A \subset \mathbb{R}^2$. Despite the fact that such a function need not be defined by a simple formula (we only have to have a rule for finding the values $f(x, y)$ at all point $(x, y) \in A$, a rule that is not ambiguous and produces one value $f(x, y)$), we will usually consider functions that are specified by a simple formula.

Examples are

- (i) $f(x, y) = 1 + 2x - 3y$
- (ii) $f(x, y) = \frac{x + y}{1 - x - y^2}$
- (iii) $f(x, y) = \sqrt{1 - x^2 - y^2}$
- (iv) $f(x, y) = \cos x \cos y$

Technically we should say not only what the rule is for finding the values $f(x, y)$, we should also say what the domain is. However, what we usually do is take the domain to be either the whole of \mathbb{R}^2 , or the largest subset of \mathbb{R}^2 where the given formula makes sense. That means excluding from the domain points (x, y) where we would be dividing by zero in the formula, or points where we would end up taking the square root of something negative, or points where we would be taking the logarithm \ln of something that is not strictly positive.

In the example $f(x, y) = \sqrt{1 - x^2 - y^2}$, we need to restrict the domain to those (x, y) where $1 - x^2 - y^2 \geq 0$. This set is reasonably easy to describe, if we rearrange the inequality as $x^2 + y^2 \leq 1$. It means

$$\{(x, y) \in \mathbb{R}^2 : (\text{dist}((x, y), (0, 0)))^2 \leq 1\}$$

and that is the same as the points of distance at most 1 from the origin in the plane \mathbb{R}^2 . This set is then the ‘closed unit disc’, the unit circle together with its inside disc.

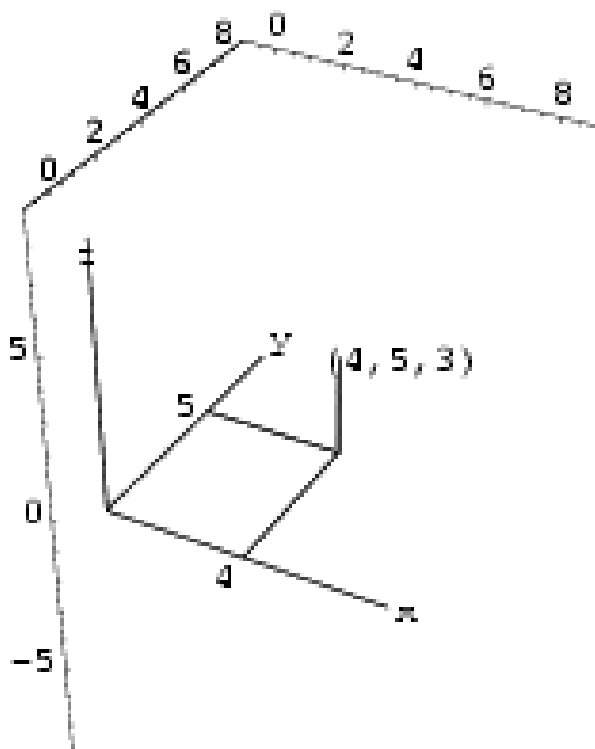
Though we have used $f(x, y)$ above, we could use any other two names for the two variables. For example $f(x_1, x_2)$ can be handier at times.

2.7 Graphs of functions of two variables. For a function of two variables, one way to picture it is via its graph. When you remember that graphs of a function of one variable (ordinary scalar

valued function of one variable) are in two dimensions (the plane), you should not be surprised that a graph of a function $f(x, y)$ will live in 3-dimensional space \mathbb{R}^3 .

We already need a plan (think of the horizontal plane or the floor) to picture the points (x, y) in the domain of the function, and then what we do for the graph is plot one point (x, y, z) with $z = f(x, y)$ the height or altitude of the point above (x, y) in the horizontal plane.

For example if $f(4, 5) = 3$, one of the points we would plot on the graph would be the point $(4, 5, 3)$ that is 3 units above the point $(4, 5)$ in the horizontal plane:



When we plot all the points (x, y, z) with $z = f(x, y)$ we end up with one point plotted above each point (x, y) in the horizontal plane and then what we get is a surface in space.

For example in the case $f(x, y) = 1 + 2x - 3y$ (a very simple formula as the terms are all constant or linear) the graph $z = f(x, y)$ is

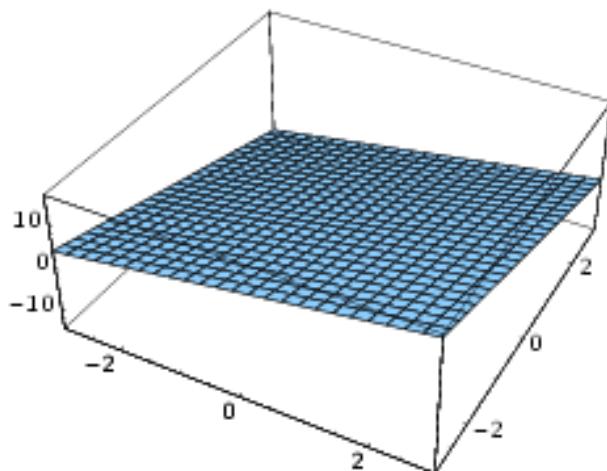
$$z = 1 + 2x - 3y$$

and we can recognise that as the equation of a plane

$$-2x + 3y + z = 1$$

which has $(-2, 3, 1) = -2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$ as a normal vector. Perhaps a way to draw it is to notice that the plane crosses the axes in $(-1/2, 0, 0)$, $(0, 1/3, 0)$ and $(0, 0, 1)$. There is just one plane

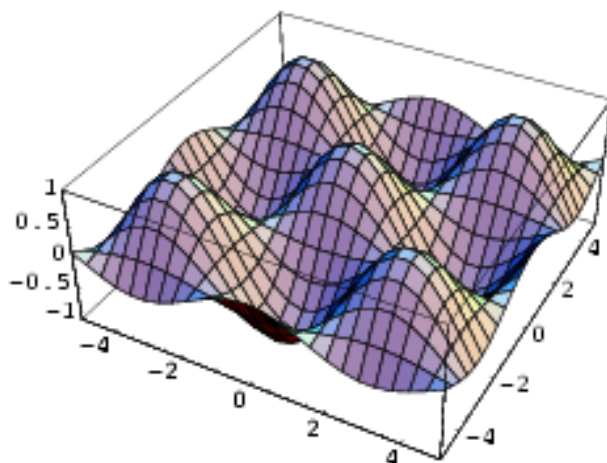
through 3 points, unless the points are in a line. Here is a plot of that plane drawn by the program Mathematica:



In the case $f(x, y) = \cos x \cos y$, we can try to piece together what the graph $z = f(x, y)$ looks like by looking at sections of the graph where x or y is constant. For example if we freeze $y = 0$, we are looking at the section of the whole graph above the x -axis, where we see $z = f(x, 0) = \cos x$, a standard cosine wave starting at $z = 1$ when $x = 0$.

If we look at what happens when $x = 0$, which is along the y -axis, we get $z = f(0, y) = \cos y$ — the same shape again. If, instead of taking a section with $x = 0$ we take a section where $x = x_0$ is some other constant we get the graph $z = \cos x_0 \cos y$. This is a modified version of the regular cosine wave $z = \cos y$, which now oscillates between $z = \cos x_0$ and $z = -\cos x_0$. Looking at sections in the perpendicular direction, where $y = y_0$ is constant we get a similar shape $z = f(x, y_0) = \cos x \cos y_0 = (\cos y_0) \cos x$.

The whole graph $z = f(x, y)$ is a sort of landscape. If we think of the x -axis as pointing East and the y -axis as pointing North (along the horizontal floor), we have a landscape that always looks like a cosine wave if we follow it either East-West or North-South. Perhaps it takes some intuition to see what it is really like and so here is a picture drawn by Mathematica.



You can see peaks and valleys. The highest points happen when $\cos x = \sin y = 1$, or when $\cos x = \sin y = -1$, and there we get $z = 1$. The lowest points happen when $\cos x = 1$ and $\sin y = -1$, or when $\cos x = -1$ and $\sin y = 1$. At these lowest points we get $z = -1$.

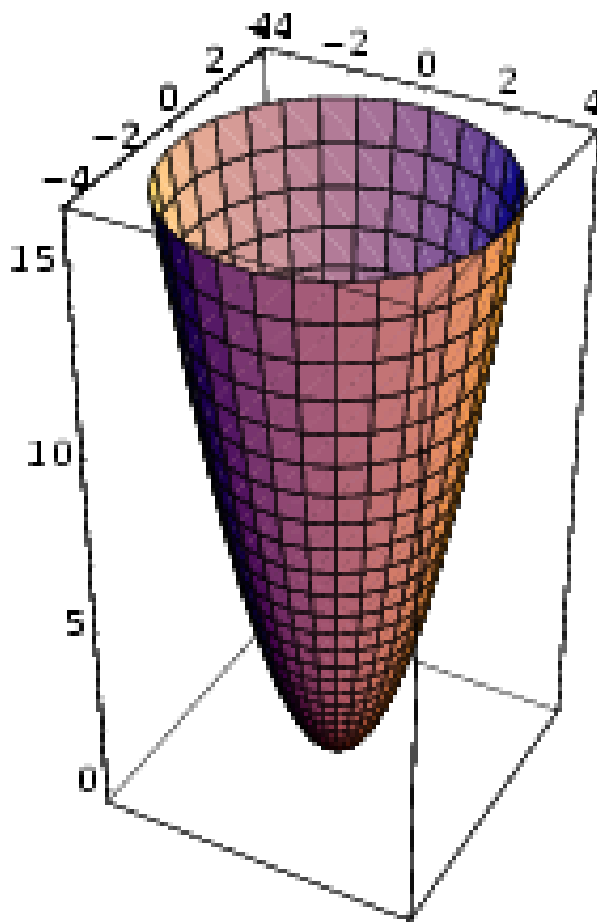
This may be a good place to explain that the altitude $z = f(x, y)$ of the points on a graph can be negative as well as positive. Negative altitude points happen when the graph dips below the floor level.

Next we look at $z = x^2 + y^2$, a somewhat easier example to figure out. This is the graph for $f(x, y) = x^2 + y^2$, and if you look to see what happens when $y = 0$ we get $z = f(x, 0) = x^2 + 0$, or $z = x^2$. This is a familiar parabola graph of x^2 . For the section along the y -axis, where $x = 0$ we get $z = f(0, y) = y^2$ and this is again the same shape.

If you take a section of this graph in some other direction from the origin, you get still the same shape because

$$f(x, y) = x^2 + y^2 = (\text{dist}((x, y), (0, 0)))^2.$$

When you piece this information together, you can see that this is what the graph looks like.



This is a graph that is rotationally symmetric around the z -axis and looks like a standard parabola (x^2 graph) when you take any vertical section through the origin. It is called a paraboloid

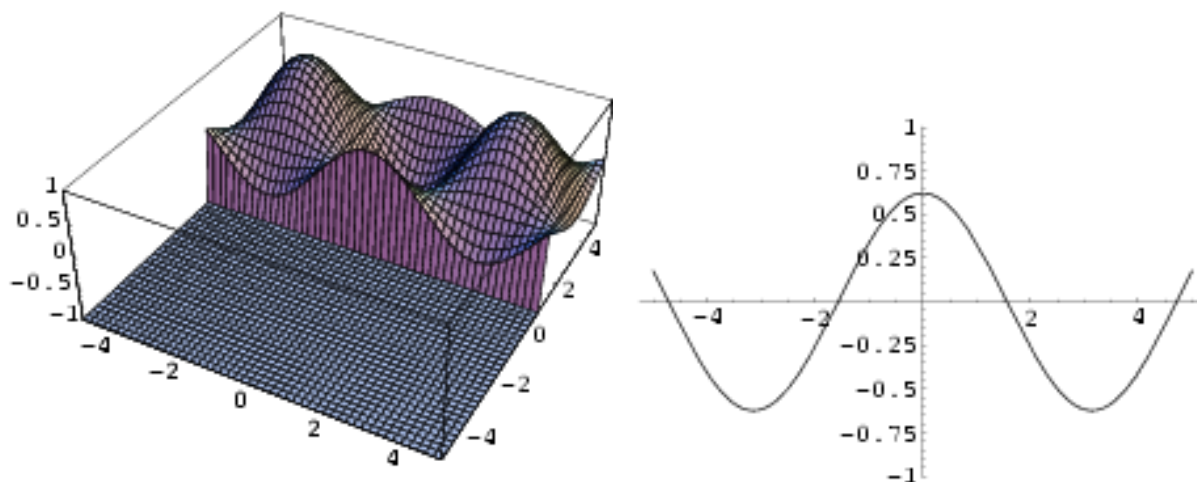
and in fact this is the shape for a reflector in a standard torch (or car head-lamps when they used to be just round). So it is a shape that has some applications.

2.8 Partial derivatives. We move on now to calculus for functions of two variables. There are a number of different concepts to absorb, but the simplest to explain is the idea of a partial derivative.

Start with a function $z = f(x, y)$ (which we can visualise in terms of its graph) and a point (x_0, y_0) . To get the partial derivative with respect to x of $f(x, y)$ at the point (x_0, y_0) we consider the function of a single variable x that we get by fixing y to be y_0 . That is the function $x \mapsto f(x, y_0)$, or $z = f(x, y_0)$.

In terms of the graph $z = f(x, y)$ of the whole function of two variables, we can see the values $z = f(x, y_0)$ as those values along the line $y = y_0$ parallel to the x -axis in the (horizontal) x - y -plane. That means that the graph of the function of one variable $z = f(x, y_0)$ can be got from the whole graph of the function of two variables by taking a section at $y = y_0$.

Here is an attempt at a picture showing what the section of $z = \cos x \cos y$ at $y = 0.9$ looks like:



One is a three dimensional picture cutting away the part of the graph $z = f(x, y)$ where $y < 0.9$ and the other is a simple graph of the profile.

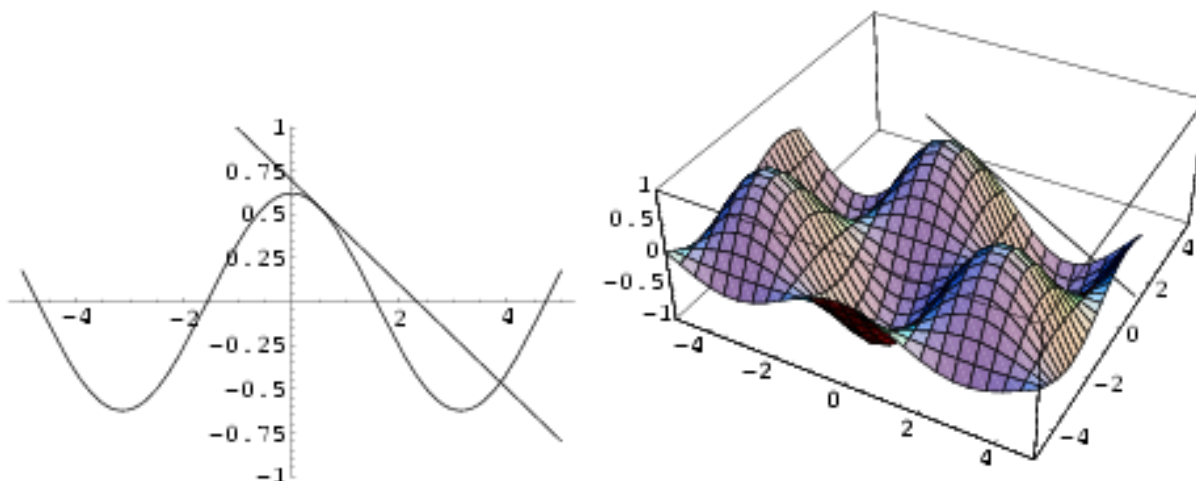
What we do now is take the derivative with respect to x of this $f(x, y_0)$ and evaluate it at $x = x_0$. This gives the partial derivative at (x_0, y_0) , the partial derivative with respect to x . While we denote regular derivatives either by a d/dx or a prime notation, we denote the partial derivative by $\partial/\partial x$. So the partial derivative evaluated at (x_0, y_0) is denoted

$$\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)}, \frac{\partial z}{\partial x} \Big|_{(x_0, y_0)}, \text{ or sometimes } f_x(x_0, y_0).$$

(We will mostly use the $\partial f/\partial x$ notation rather than the f_x notation, but the f_x notation is also used quite often.) We could summarise the definition as

$$\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} = \frac{d}{dx} \Big|_{x=x_0} f(x, y_0)$$

If we want to see graphically what we are getting, recall that the ordinary derivative (of a function of one variable) is the slope of the tangent line to the graph at the point. So if for example, we were taking $f(x, y) = \cos x \cos y$, $(x_0, y_0) = (0.5, 0.9)$ we would be calculating the slope of this line



We can calculate this slope (or this derivative) exactly:

$$\begin{aligned}
 f(x, y_0) &= \cos x \cos y_0 \\
 &= \cos x \cos 0.9 \\
 \frac{d}{dx} f(x, y_0) &= -\sin x \sin y_0 \\
 &= -\sin x \sin 0.9 \\
 \frac{d}{dx} \big|_{x=x_0} f(x, y_0) &= -\sin x_0 \cos y_0 \\
 &= -\sin 0.5 \cos 0.9 = -0.298016 \\
 \frac{\partial f}{\partial x} \big|_{(x_0, y_0)} &= -0.298016
 \end{aligned}$$

Looking at the calculation, you can see that what we do is replace y by the constant y_0 and then take the derivative d/dx with respect to x (and finally put $x = x_0$). We usually do it in a slightly quicker way. When taking the partial $\partial/\partial x$ with respect to x , we treat the variable y as a constant. (We don't have to go so far as to replace y by a specific number y_0 .) If we repeat the above calculation this way, it would look like this:

$$\begin{aligned}
 f(x, y) &= \cos x \cos y \\
 \frac{\partial f}{\partial x} &= -\sin x \cos y \\
 \frac{\partial f}{\partial x} \big|_{(x_0, y_0)} &= -\sin x_0 \cos y_0 = -\sin 0.5 \cos 0.9 \\
 &= -0.298016.
 \end{aligned}$$

Summary: To get the partial derivative with respect to x , take the derivative with respect to x while treating the variable y as though it was a constant.

Now partial derivatives with respect to y are similar, except we freeze $x = x_0$ and differentiate with respect to a variable y this time.

$$\frac{\partial f}{\partial y} \big|_{(x_0, y_0)} = \frac{d}{dy} \big|_{y=y_0} f(x_0, y)$$

Or, we can summarise by saying: To get the partial derivative with respect to y , take the derivative with respect to y treating the variable x as though it was a constant.

Graphically it means taking a section through our graph $z = f(x, y)$ in the x -direction, and then finding the slope of that.

If we take our example again $f(x, y) = \cos x \cos y$, $(x_0, y_0) = (0.5, 0.9)$, and find the partial with respect to y this time, we find

$$\begin{aligned} f(x, y) &= \cos x \cos y \\ \frac{\partial f}{\partial y} &= \cos x (-\sin y) \\ &= -\cos x \sin y \\ \frac{\partial f}{\partial y} \big|_{(x_0, y_0)} &= -\cos x_0 \sin y_0 = -\cos 0.5 \sin 0.9 \\ &= -0.687434. \end{aligned}$$

2.9 Examples. It is necessary to become accustomed to taking partial derivatives. The rules are as for derivatives of functions of one variable, and so it is a question of getting practice in these again.

You need to remember the rules — chain rule, product rule, perhaps the quotient rule (or else know how to treat $\frac{u}{v} = uv^{-1}$ as the product of u times the reciprocal of v), and the basic differentiation formulae for $\cos t$, $\sin t$, e^t , $\ln t$, $\tan t$, $\sec t$ and so forth.

Exercise: Look at the exercises in Thomas' Calculus (§14.3 the 11th edition, or §11.3 in the 10th) entitled 'Calculating first order Partial derivatives' and try them for yourself. The book has the answers to the odd numbered ones so you can see how you are doing.

2.10 Directional derivatives. With partial derivatives we have concentrated on (vertical) sections in the directions parallel to the two axes (the x -axis for $\partial/\partial x$ and the y -axis for $\partial/\partial y$), but there seems to be no compelling reason to take only the slope of our graph $z = f(x, y)$ in these two directions.

Why should we not consider all possible directions through (x_0, y_0) ? The slope we get is called a directional derivative.

To explain it, we should find a good way to describe a direction. We know vectors in the plane have a direction and a magnitude. As we are only interested in a direction, we can stick to unit vectors (magnitude = 1) \mathbf{u} to describe a direction. Going through (x_0, y_0) in the direction \mathbf{u} we have a line and we know how to describe this line with parametric equations:

$$(x, y) = (x_0, y_0) + t\mathbf{u}$$

At $t = 0$ we have our point (x_0, y_0) and we are heading in the direction specified by \mathbf{u} . If we now look at the values of $f(x, y)$ along this line, we get the function

$$t \mapsto f((x_0, y_0) + t\mathbf{u})$$

of the parameter (single variable) t . The graph of this function of t is what we would get by taking a (vertical) section through the graph $z = f(x, y)$. It is the section through (x_0, y_0) in the direction \mathbf{u} . We are going to take the slope of this section at the point (x_0, y_0) we are concentrating on. We can see that the slope is the same¹ as the slope of the function of t , that is the same as the derivative with respect to t at $t = 0$:

$$\frac{d}{dt} \Big|_{t=0} f((x_0, y_0) + t\mathbf{u})$$

We *define* the directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ of f in the direction \mathbf{u} at (x_0, y_0) to be

$$D_{\mathbf{u}}f(x_0, y_0) = \frac{d}{dt} \Big|_{t=0} f((x_0, y_0) + t\mathbf{u})$$

We can notice that if we take $\mathbf{u} = \mathbf{i} = (1, 0)$ = the unit vector in the direction of the positive x -axis, then

$$\begin{aligned} D_{\mathbf{u}}f(x_0, y_0) &= D_{\mathbf{i}}f(x_0, y_0) \\ &= \frac{d}{dt} \Big|_{t=0} f((x_0, y_0) + t\mathbf{i}) \\ &= \frac{d}{dt} \Big|_{t=0} f((x_0 + t, y_0)) \\ &= \frac{d}{dx} \Big|_{x=x_0} f(x, y_0) \\ &= \frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \end{aligned}$$

Similarly for $\mathbf{u} = \mathbf{j} = (0, 1)$ the directional derivative $D_{\mathbf{j}}f$ is the same as $\partial f / \partial y$.

We can compute what $D_{\mathbf{u}}f(x_0, y_0)$ for the example we have used above, that is $f(x, y) = \cos x \cos y$ and $(x_0, y_0) = (0.5, 0.9)$. We'll work it out for an arbitrary unit vector $\mathbf{u} = (u_1, u_2)$.

$$\begin{aligned} D_{\mathbf{u}}f(x_0, y_0) &= \frac{d}{dt} \Big|_{t=0} f((x_0, y_0) + t\mathbf{u}) \\ &= \frac{d}{dt} \Big|_{t=0} f(x_0 + tu_1, y_0 + tu_2) \\ &= \frac{d}{dt} \cos(x_0 + tu_1) \cos(y_0 + tu_2) \\ &= (-u_1 \sin(x_0 + tu_1) \cos(y_0 + tu_2) - \cos(x_0 + tu_1) u_2 \sin(y_0 + tu_2)) \Big|_{t=0} \\ &= -u_1 \sin x_0 \cos y_0 - u_2 \cos x_0 \sin y_0 \\ &= -\sin x_0 \cos y_0 u_1 - \cos x_0 \sin y_0 u_2 \\ &= -0.298016u_1 - 0.687434u_2. \end{aligned}$$

¹For this to be true we are relying on the fact that \mathbf{u} is a unit vector — if we allowed non unit vectors, the scale along the t axis would not correspond to the distance scale in the horizontal (x, y) plane, and this would affect the calculation of the slope.

You'll notice that the same numbers are coming in as we had for the partial derivatives. Of course this was just one example of a very simple kind, and the fact that the directional derivative involves the same numbers we got from the partial derivatives might be a fluke.

If you try more examples though, you will begin to see that we seem to always have this

$$D_{\mathbf{u}}f(x_0, y_0) = \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) u_1 + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) u_2 \quad (1)$$

Or we might be better to use the other notation for partials

$$D_{\mathbf{u}}f(x_0, y_0) = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2$$

or to just write the formula at an arbitrary (x, y) (without trying to write down that both sides have to be evaluated at the same point):

$$D_{\mathbf{u}}f = \frac{\partial f}{\partial x}u_1 + \frac{\partial f}{\partial y}u_2.$$

Anyhow, it turns out that this formula is usually true. Not quite always, but usually in examples we can easily compute with. We will explain a little more carefully later when it is a valid formula.

The upshot of all this is that the values of the partial derivatives contain the information needed to find all the directional derivatives.

2.11 Tangent plane to a graph. In order to explain what it is that makes the formula for directional derivatives given above true, and the circumstances where we can be sure it is true, we explain about the tangent plane to a graph $z = f(x, y)$ at a point (x_0, y_0, z_0) on the graph.

Just as in one variable there are some graphs with no tangent line, or functions that are not differentiable at certain points, there are cases of graphs $z = f(x, y)$ that do not have a tangent plane. Before we can make that more clear, we will need to say what we mean by a tangent plane.

What we will do first is work out what the equation for the tangent plane should be, assuming there is a tangent plane. Then we will be able to say what the definition of a tangent plane is.

So we look for the tangent plane to a graph $z = f(x, y)$ at a point (x_0, y_0, z_0) on the graph. One simple thing that we need in practice is that if the point is on the graph then $z_0 = f(x_0, y_0)$. (We use this to work out the z -coordinate if given the x - and y -coordinates.)

Then we consider the sections of the graph in the x - and y -directions, the same sections we looked at when we were describing partial derivatives. Rather than considering the graphs $z = f(x, y_0)$ (where we fix $y = y_0$) and $z = f(x_0, y)$ (where we fix $x = x_0$) as graphs in a vertical plane we look at them as curves in space. We can parametrise them as

$$\begin{aligned} x &= t \\ y &= y_0 \\ z &= f(t, y_0) \end{aligned}$$

and

$$\begin{aligned}x &= x_0 \\y &= t \\z &= f(x_0, t_0).\end{aligned}$$

Look at the first of these

$$(x, y, z) = (t, y_0, f(t, y_0))$$

which passes by our point (x_0, y_0, z_0) at $t = x_0$. It will have a tangent vector at that point

$$\frac{d}{dt} \Big|_{t=x_0} (t, y_0, f(t, y_0)) = \left(1, 0, \frac{d}{dt}f(t, y_0)\right) \Big|_{t=x_0} = \left(1, 0, \frac{\partial f}{\partial x} \Big|_{(x_0, y_0)}\right)$$

Whatever the tangent plane to the graph $z = f(x, y)$ is, it seems reasonable to believe that this tangent vector should be in the plane (or parallel to it).

Similarly working with the curve we get along the y -direction, we have

$$(x, y, z) = (x_0, t, f(x_0, t))$$

and the tangent vector to that (at $t = y_0$) will come out as

$$\frac{d}{dt} \Big|_{t=y_0} (x_0, t, f(x_0, t)) = \left(0, 1, \frac{d}{dt}f(x_0, t)\right) \Big|_{t=y_0} = \left(0, 1, \frac{\partial f}{\partial y} \Big|_{(x_0, y_0)}\right).$$

This vector should be in the plane also, for the same reason.

So our tangent plane should be parallel to both the vectors

$$\mathbf{A} = (1, 0, f_x(x_0, y_0)) \text{ and } \mathbf{B} = (0, 1, f_y(x_0, y_0)).$$

From what we did in Example 2.4 above we know then that $\mathbf{A} \times \mathbf{B}$ should be normal to our tangent plane. We can work that out:

$$\begin{aligned}\mathbf{A} \times \mathbf{B} &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x(x_0, y_0) \\ 0 & 1 & f_y(x_0, y_0) \end{pmatrix} \\&= \mathbf{i} \det \begin{pmatrix} 0 & f_x(x_0, y_0) \\ 1 & f_y(x_0, y_0) \end{pmatrix} - \mathbf{j} \det \begin{pmatrix} 1 & f_x(x_0, y_0) \\ 0 & f_y(x_0, y_0) \end{pmatrix} + \mathbf{k} \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\&= -f_x(x_0, y_0)\mathbf{i} - f_y(x_0, y_0)\mathbf{j} + \mathbf{k} \\&= (-f_x(x_0, y_0), -f_y(x_0, y_0), 1)\end{aligned}$$

We can then write down that the equation of the plane through (x_0, y_0, z_0) perpendicular to that vector should be

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

Rearranging it to have fewer minus signs, we get

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

Recalling that $z_0 = f(x_0, y_0)$, we find that the tangent plane should have this equation

$$z = f(x_0, y_0) + \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0)$$

(if there is a tangent plane).

2.12 Linear approximation. We turn now to the definition of what it means for $z = f(x, y)$ to have a tangent plane at (x_0, y_0, z_0) . The idea is fairly simple — the graph $z = f(x, y)$ should be close to the graph of the plane near the point (x_0, y_0) .

We should be able to say that there is a plane

$$z = z_0 + a_1(x - x_0) + a_2(y - y_0)$$

(or graph of a function with only constant and linear terms) through (x_0, y_0, z_0) so that it is true that

$$f(x, y) \cong z_0 + a_1(x - x_0) + a_2(y - y_0)$$

is a good approximation when we keep (x, y) not too far from (x_0, y_0) . Unless we are a bit more precise about what good means, this will still be ambiguous. We want to insist on something that can't be true of two different planes and is still often achievable.

What we do is insist that the error in the above approximation should be small compared to the distance from (x, y) to (x_0, y_0) (when that distance is itself small).

We make the definition that a function $z = f(x, y)$ is *differentiable* at (x_0, y_0) if it is possible to find constants a_1 and a_2 so that

$$\lim_{(x, y) \rightarrow (x_0, y_0)} \frac{f(x, y) - (f(x_0, y_0) + a_1(x - x_0) + a_2(y - y_0))}{\text{dist}((x, y), (x_0, y_0))} = 0.$$

We could rewrite that

$$\lim_{(x, y) \rightarrow (x_0, y_0)} \frac{f(x, y) - (f(x_0, y_0) + a_1(x - x_0) + a_2(y - y_0))}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0. \quad (2)$$

Sometimes the term *total derivative* is used for this rather than derivative.

We are being a bit cavalier here in bringing in limits of functions of two variables with no explanation. The book by Thomas has a lot more than we are going to do about this. One thing to point out is that when we let (x, y) approach (x_0, y_0) , we never actually take (x, y) to be equal to (x_0, y_0) . That means we don't end up having to worry about interpreting $0/0$ in the above limit. Another important thing is that (x, y) does not just approach (x_0, y_0) from one direction, but from all directions at once. The interpretation of the limit being 0 is that the thing we are taking the limit of will be close to 0 as long as we take (x, y) within some small distance of (x_0, y_0) . But we must not have to make the small distance be 0, and we should never consider $(x, y) = (x_0, y_0)$.

We claim now that the number a_1 in the definition must always be the partial derivative with respect to x at (x_0, y_0) (and the number a_2 must be the partial with respect to y). For this we go back to first principles, and write the partial with respect to x as a limit

$$\frac{\partial f}{\partial x} \big|_{(x_0, y_0)} = \frac{d}{dx} \big|_{x=x_0} f(x, y_0) = \lim_{x \rightarrow x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0}$$

What we are going to prove is that, if we know (2), then the partial derivative with respect to x must exist at (x_0, y_0) and must be the number a_1 . We do this by looking at the difference

$$\begin{aligned} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0} - a_1 &= \frac{f(x, y_0) - f(x_0, y_0) - a_1(x - x_0)}{x - x_0} \\ &= \frac{f(x, y_0) - (f(x_0, y_0) + a_1(x - x_0))}{x - x_0} \end{aligned}$$

and its absolute value

$$\left| \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0} - a_1 \right| = \frac{|f(x, y_0) - (f(x_0, y_0) + a_1(x - x_0))|}{|x - x_0|}.$$

From (2) we know that this is small as long as

$$\text{dist}((x, y_0), (x_0, y_0))$$

is small, which is the same as

$$|x - x_0| \text{ small.}$$

In this way we prove that

$$\lim_{x \rightarrow x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0} = a_1$$

must be true if (2) is true. That is

$$\frac{\partial f}{\partial x} \big|_{(x_0, y_0)} = a_1$$

Similarly, it is possible to show that if (2) is true, then

$$\frac{\partial f}{\partial y} \big|_{(x_0, y_0)} = a_2$$

and in fact that for any unit vector \mathbf{u}

$$D_{\mathbf{u}}f(x_0, y_0) = a_1u_1 + a_2u_2$$

must be true if (2) holds.

This justifies the equation (1) — **if** we know that $f(x, y)$ has a derivative at (x_0, y_0) . In fact, most functions that we can easily write down do have derivatives. At the worst they usually can have some bad points where things go wrong, points (x_0, y_0) where there is no derivative.

To make a precise statement, here is the theorem (but we will not go any way towards proving it). If the partial derivatives $f_x(x, y) = \partial f / \partial x$ and $f_y(x, y) = \partial f / \partial y$ exist at all points near (x_0, y_0) and we have

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f_x(x, y) = f_x(x_0, y_0) \text{ and } \lim_{(x,y) \rightarrow (x_0,y_0)} f_y(x, y) = f_y(x_0, y_0)$$

then f is differentiable at (x_0, y_0) .

With one more bit of terminology, we can state this more tidily. A function $f(x, y)$ is called *continuous* at the point (x_0, y_0) if it is true that

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y) = f(x_0, y_0).$$

We won't discuss this much. (There is some discussion in the book by Thomas.) The summary is that this is usually true, apart from maybe one or two bad points, of functions that can be easily written down by a formula. In the practical real world, there can be problems because functions that have to be considered arise from the real world rather than from nice tidy formulae. In these cases, we could be faced with discontinuous functions.

The theorem stated above can be restated like this

Theorem. If a function $z = f(x, y)$ has partial derivatives $\partial f / \partial x$ and $\partial f / \partial y$ that are continuous functions at (x_0, y_0) , then f has a derivative at (x_0, y_0) .

This means (2) holds with

$$a_1 = \frac{\partial f}{\partial x} \Big|_{(x_0,y_0)} \text{ and } a_2 = \frac{\partial f}{\partial y} \Big|_{(x_0,y_0)}$$

The reason this is useful is that we can usually see that we are in the situation where the partial derivatives are continuous. You can compute partial derivatives easily and then looking at the formulae you get, you can usually see that you are in the situation where they give you continuous functions.

As we are not going to go into this in any detail, we will just summarise by saying there are theorems that justify the claim that most of the functions we encounter have derivatives, and then the directional derivative formula (1) is valid.

Restating (2) in a less precise way, and using what we know about what a_1 and a_2 have to be we get the *linear approximation formula*

$$f(x, y) \cong f(x_0, y_0) + \left(\frac{\partial f}{\partial x} \Big|_{(x_0,y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0,y_0)} \right) (y - y_0)$$

which is valid for (x, y) close to (x_0, y_0) .

There is an example of working this out with numerical values for everything on the solutions to Tutorial 7 (question 2). But the more important fact is that linear approximation is a way to

tackle complicated problems by reducing them to linear ones. We can solve most linear problems. The theory of linear problems (linear algebra) is well enough understood that we know how to do most things. The difficulty is that many of the interesting problems we encounter in the real world of applications are not linear. With linear approximation we can tackle these problems by replacing them by linear problems (which we can usually solve pretty easily). However, we are then left with the question of how to relate the solution to the approximate problem to the actual nonlinear problem.

2.13 The gradient vector. We define the gradient vector at (x_0, y_0) of a function (x, y) of two variables as

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

The symbol ∇ has a name “nabla” but we usually just pronounce it ‘gradient’. We can maybe simplify the above definition by removing the evaluation at (x_0, y_0) and writing

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

where we understand that both sides are to be evaluated at the same place.

We’ve seen that the partial derivatives are crucial numbers for several purposes and the idea of the gradient is to wrap up both numbers in one (vector) quantity. It is perhaps a short step, but it is quite a useful step.

2.14 The gradient vector and directional derivatives. We can restate the formula (1) using the gradient vector as

$$D_{\mathbf{u}}f(x_0, y_0) = (\nabla f|_{(x_0, y_0)}) \cdot \mathbf{u}$$

Remembering how dot products work, we can write this as

$$D_{\mathbf{u}}f(x_0, y_0) = \|\nabla f|_{(x_0, y_0)}\| \|\mathbf{u}\| \cos \theta$$

where θ is the angle between the gradient vector and the direction \mathbf{u} . Since $\|\mathbf{u}\| = 1$, we have

$$D_{\mathbf{u}}f(x_0, y_0) = \|\nabla f|_{(x_0, y_0)}\| \cos \theta$$

Since $-1 \leq \cos \theta \leq 1$, we can then say that

$$-\|\nabla f|_{(x_0, y_0)}\| \leq D_{\mathbf{u}}f(x_0, y_0) \leq \|\nabla f|_{(x_0, y_0)}\|$$

Moreover, we can say that the largest directional derivative is the magnitude of the gradient vector and it is achieved only when the direction \mathbf{u} is the unit vector in the direction of the gradient vector. That is the largest directional derivative is when

$$\mathbf{u} = \frac{\nabla f|_{(x_0, y_0)}}{\|\nabla f|_{(x_0, y_0)}\|}$$

The smallest possible directional derivative is minus the length of the gradient vector and that happens only when \mathbf{u} is the unit vector in the opposite direction to the gradient.

(All this supposes that the gradient vector is not the zero vector, which can happen. In that case all the directional derivatives are 0.)

2.15 The Chain Rule for partial derivatives. We can now state one version of the chain rule that involves functions of more than one variable.

First a quick refresher on the chain rule in one variable. It says that if $y = f(x)$ and $x = x(t)$ so that in the end we can regard y as a function of t by $y = f(x(t))$, then

$$\frac{dy}{dt} = \frac{dy}{dx} \frac{dx}{dt}.$$

In other words it is ok to cancel the dx 's in this Leibniz notation for derivatives.

For a function $z = f(x, y)$ of two variables where each of x and y is itself dependent on t by $x = x(t)$ and $y = y(t)$, we can ask how the derivative dz/dt is related to the derivatives of the functions that make up the composition $z = f(x(t), y(t))$. The formula is

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

So the simple rule of “cancelling the dx 's” becomes less obvious for partial derivatives. Somehow we end up with only one fraction dz/dt when we ‘cancel’ both the partials ∂x and ∂y .

In vector form we can express the formula as

$$\frac{dz}{dt} = (\nabla f) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt} \right)$$

So we can see the gradient multiplied (or dot producted with) the derivative vector (or velocity vector) of the parametric curve

$$(x, y) = (x(t), y(t))$$

We won't prove this formula, but the way it is usually proved relies on linear approximation. Fix on a value $t = t_0$ where we want to prove the chain rule formula. Then take $(x_0, y_0) = (x(t_0), y(t_0))$ in the linear approximation formula

$$f(x, y) \cong f(x_0, y_0) + \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0).$$

Into this we put the linear approximation formulae in one variable for $x(t)$ and $y(t)$ around $t = t_0$. These say

$$\begin{aligned} x(t) &\cong x(t_0) + x'(t_0)(t - t_0) \\ y(t) &\cong y(t_0) + y'(t_0)(t - t_0) \end{aligned}$$

(or that the graphs $x = x(t)$ and $y = y(t)$ are close to the graphs of their tangent lines at t_0 , if we keep t near t_0 .)

Combining the three linear approximations we get

$$f(x(t), y(t)) \cong f(x_0, y_0) + \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) x'(t_0)(t - t_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) y'(t_0)(t - t_0)$$

and this leads to a proof that

$$\frac{d}{dt} f(x(t), y(t)) \Big|_{t=t_0} = \lim_{t \rightarrow t_0} \frac{f(x(t), y(t)) - f(x(t_0), y(t_0))}{t - t_0}$$

has the value it is supposed to have in the chain rule formula.

We will not go into details.

2.16 Level curves. We have concentrated so far on visualising function $z = f(x, y)$ by their graphs, which are surfaces (or look like landscapes) in space.

There is another way which we will now consider. It is to draw (on a two dimensional plane) the curves

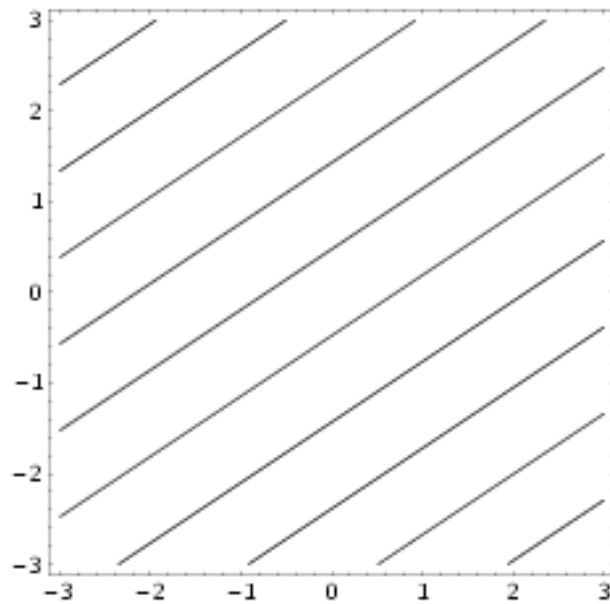
$$f(x, y) = c$$

for different values of c . These are called the *level curves* for the function $f(x, y)$.

There are everyday examples where we see these curves. On the weather chart on the TV or in the newspaper you see lines (called isobars) where the pressure is constant. So these are level curves for the function $f(x, y) = \text{atmospheric pressure at point } (x, y) \text{ on the earth}$. [Perhaps there is a small problem as the earth is not a plane.] We also sometimes see charts of the temperature today or tomorrow, usually coloured (so that red zones are hot, blue zones are cold). The dividing curves between the different colours are curves where the temperature is constant (30 degrees, 25 degrees, or whatever). So they are level curves for the temperature function.

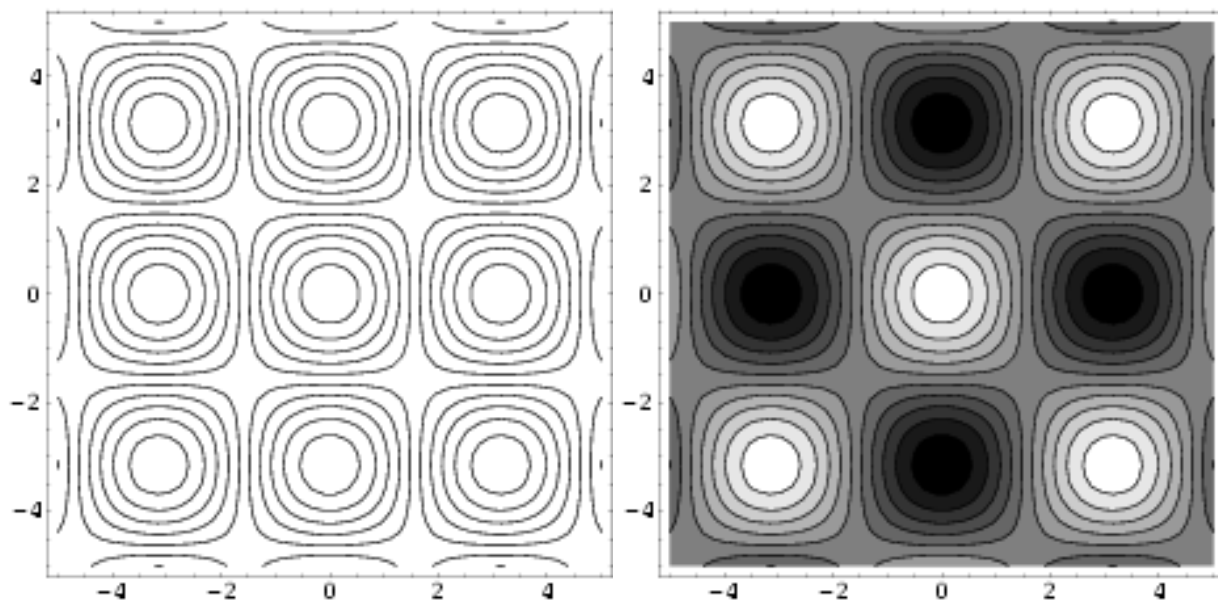
An example that is maybe better for us are the contour lines we see on certain maps. There will be many such contour lines around mountains or hills, perhaps marked with 100 metres, 200 metres and so on. They are lines on the map of tracks you could in principle follow that would be neither uphill nor downhill. They are lines of constant altitude, or level curves for the function $f(x, y) = \text{altitude}$. Off shore they can represent negative altitudes (or depths of the sea bed). This is exactly how our level curves for functions $f(x, y)$ correspond to the graph $z = f(x, y)$ of a function of two variables.

We can maybe look to some of our examples to see how these level curves work out. Here is the example $z = 1 + 2x - 3y$ we plotted first done with contours:



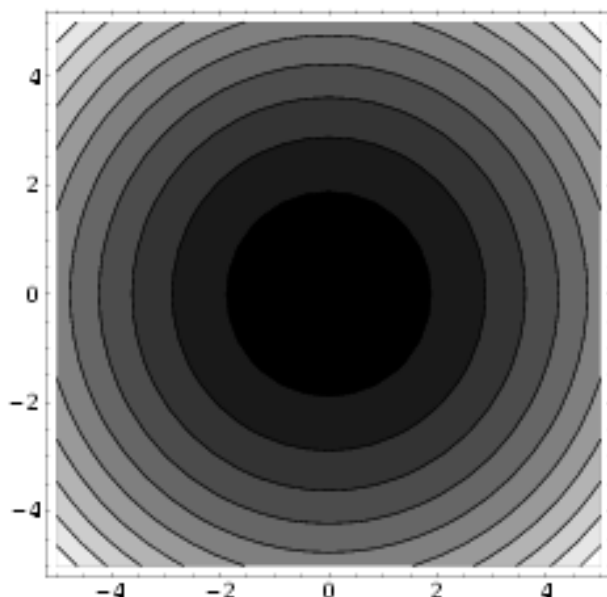
It is maybe not so useful without indicating what the values of c is on each line. We could see by hand that $1 + 2x - 3y = c$ is a line $y = (2/3)x + (c - 1)/3$ of slope $2/3$ and that is what we see in the picture — various lines of that slope.

Here is the picture for $z = \cos x \cos y$, once with just the level curves, and then with shading.



The lighter shaded regions (like near the origin $(x, y) = (0, 0)$) are higher values.

For the example $z = x^2 + y^2$, it is quite easy to see that the level curves $x^2 + y^2 = c$ are circles (of radius \sqrt{c}) around the origin if $c > 0$. For $c = 0$ the level ‘curve’ $x^2 + y^2 = 0$ is just the one point $(0, 0)$, and for $c < 0$ the level ‘curve’ is empty. Here is a (shaded) of some of the contours from the Mathematica programme:



From the fact that the level curves are circles we could immediately realise that the function is symmetrical around the origin, if we did not know that already. We can build the graph, which we saw before was a paraboloid, by lifting up the level curves $x^2 + y^2 = c$ to height c above the horizontal plane.

This may help you to see the relationship between the level curves and the graph. The map example, where they correspond to tracks you can walk along the landscape without going up or down at any stage, is also helpful.

Though a graph is maybe easier to visualise, it takes 3 dimensions. Maps with contour lines showing (say) the Mourne mountains could be harder to interpret, but they are easier to make than 3-dimensional models of the mountains.

2.17 Level curves and the gradient. Here we explain that the gradient vector at (x_0, y_0) of $f(x, y)$ will always be perpendicular to the level curve $f(x, y) = c$ that goes through the point (x_0, y_0) . That is the one where $c = f(x_0, y_0)$.

Let us start with an example, the simple linear example $z = 1 + 2x - 3y$ and the point $(x_0, y_0) = (7, 11)$. We pictured the level curves (lines) for this earlier. So in this case we are dealing with the function $f(x, y) = 1 + 2x - 3y$ and we can work out the right value of c so that the level curve $f(x, y) = c$ passes through $(x_0, y_0) = (7, 11)$. We need $f(7, 11) = c$, or $1 + 14 - 33 = c$, $c = -18$. So the level curve is $1 + 2x - 3y = -18$ or $2x - 3y = -19$. As this is a line we can get its slope by rewriting it as $2x + 19 = 3y$, or $y = (2/3)x + 19/3$, so that the slope is $2/3$. If we find the gradient of f we get

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = (2, -3).$$

So the slope of the gradient vector is $-3/2$ and the product of this with the slope of the level curve (line) is $(-3/2)(2/3) = -1$. This shows they are perpendicular (in this example).

We take now another example $f(x, y) = x^2 + y^2$ and the point $(4, -1)$. We saw this example before and we know the level curves are concentric circles centered at the origin. So the tangent lines are perpendicular to the radius. In particular the tangent line to the circle through $(4, -1)$ is perpendicular to the vector $(4, -1)$. But if we work out the gradient vector we get

$$\begin{aligned}\nabla f|_{(4,-1)} &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) |_{(4,-1)} \\ &= (2x, 2y) |_{(4,-1)} \\ &= (8, -2).\end{aligned}$$

As this gradient vector is a multiple of $(4, -1)$ we see that it is indeed perpendicular to the tangent line (to the level curve, or level circle in this example).

We've seen that the claimed fact that the gradient is always perpendicular to the tangent line to the level curve has worked out in two examples. Now we will indicate why it is always true.

The level curve $f(x, y) = c$ through a point (x_0, y_0) can be described in a different way by parametric equations

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}$$

in such a way that $(x(0), y(0)) = (x_0, y_0)$. We then have the equation

$$f(x(t), y(t)) = c$$

satisfied for all t . If we differentiate both sides with respect to t we get

$$\begin{aligned}\frac{d}{dt} f(x(t), y(t)) &= 0 \\ \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} &= 0\end{aligned}$$

by using the chain rule. When we put $t = 0$ this comes down to

$$\left(\frac{\partial f}{\partial x} |_{(x_0, y_0)}, \frac{\partial f}{\partial y} |_{(x_0, y_0)} \right) \cdot \left(\frac{dx}{dt} |_{t=0}, \frac{dy}{dt} |_{t=0} \right) = 0.$$

This means that the dot product of the gradient vector of f at the point (x_0, y_0) is perpendicular to the velocity vector to the curve.

There are some details to fill in, but this is one way to prove that the gradient is perpendicular to the level curve.

2.18 Example. Find the equation of the tangent line to

$$\cos(2x) \cos y = \frac{1}{2}$$

at the point $(\pi/8, \pi/4)$.

To do this we need the gradient vector to $f(x, y) = \cos x \cos y$ at the point $(\pi/4, \pi/4)$.

$$\begin{aligned}\nabla f|_{(\pi/8, \pi/4)} &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) |_{(\pi/8, \pi/4)} \\ &= (-2 \sin(2x) \cos y, -\cos(2x) \sin y) |_{(\pi/8, \pi/4)} \\ &= (-2(1/\sqrt{2})(1/\sqrt{2}), -(1/\sqrt{2})(1/\sqrt{2})) = (-1, -1/2)\end{aligned}$$

A quick way to write down the equation of the line through $(\pi/4, \pi/4)$ perpendicular to the vector $(-1, -1/2)$ is this:

$$(-1)\left(x - \frac{\pi}{8}\right) + \left(-\frac{1}{2}\right)\left(y - \frac{\pi}{4}\right) = 0.$$

2.19 Gradient is direction of steepest ascent. The word ‘gradient vector’ is justified by the fact that it relates to the direction of steepest ascent.

In a way we know this already because we know from 2.14 that the directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ is largest (for a fixed point (x_0, y_0)) when \mathbf{u} is the unit vector in the direction of the gradient $\nabla f|_{(x_0, y_0)}$. We also know that the largest possible value of the directional derivative is the magnitude of the gradient vector.

We can relate this to the graph $z = f(x, y)$. We can think of the graph as a landscape. Suppose we are starting from the point on the graph where $(x, y) = (x_0, y_0)$. The directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ gives us the slope we would encounter if we travelled from position (x_0, y_0) in the direction \mathbf{u} . We get the largest slope in the direction of the gradient, the smallest slope in the opposite direction.

That means that the gradient vector is the most uphill direction.

The level curve direction is the direction perpendicular to the gradient. That is what we just figured out above when we showed that the gradient vector is perpendicular to (the tangent line) to the level curve. It also makes sense that if you travel along the level curve you will encounter zero slope.

2.20 Functions of 3 variables and level surfaces. We look now at functions $f(x, y, z)$ or $f(x_1, x_2, x_3)$ of 3 variables.

A lot of what we have to say will be similar to what we did before for functions of two variables, but there will also be some differences.

The first thing that is the same (or hardly any different) is that the function is a well-defined rule associated a single value $f(x, y, z)$ to each point (x, y, z) in its domain. Usually, but not necessarily, the functions we will consider will be specified by giving a formula to find the values. Examples could be

$$\begin{aligned}f(x, y, z) &= x^2 + y^2 + z^2 \\ f(x, y, z) &= x \cos(y + z) + \frac{1}{x} \\ f(x_1, x_2, x_3) &= \frac{x_1 x_2 x_3}{\sqrt{x_1^2 + x_2^2 + x_3^2 + 1}}\end{aligned}$$

As with the earlier case of functions of two variables, a function $f(x, y, z)$ will be defined for points (x, y, z) in some domain, some subset of \mathbb{R}^3 or for points all of \mathbb{R}^3 if the domain is the whole of \mathbb{R}^3 . As with functions of two variables we will sometimes assume that the domain is as big as it can be for the formula for $f(x, y, z)$ to make sense. In the second of the three examples above, we should exclude points with $x = 0$ from the domain as we don't want to be dividing by zero. For the other two examples above we don't need any restrictions.

However, when we start to think of graphs, we find a problem. A graph of a function of one variable (a graph $y = f(x)$) we need to draw it in the plane (so we need two dimensions, one for the domain and one more for the values). Similarly when we looked at functions of two variables we needed 3 dimensions to fit in the graph $z = f(x, y)$.

A graph of a function $f(x, y, z)$ would need 4 dimensions. It already takes 3 dimensional space to keep track of the domain points (x, y, z) and we would need to find a fourth dimension to fit in the values. So we can't look at it in reality. We can do it theoretically, but then we can just work by analogy with what we are familiar with.

The level surface approach $f(x, y, z) = c$ is available. It is harder perhaps to think about than the level curves we had in the plane (for functions of two variables), but the level surfaces are in space where we can see (or visualise) them.

For example in the case $f(x, y, z) = x^2 + y^2 + z^2$, we can describe the level surfaces quite easily. For $c > 0$, the equation $f(x, y, z) = c$ means the distance squared from (x, y, z) to the origin is c . So the points (x, y, z) satisfying $f(x, y, z) = c$ are exactly the points on the sphere of radius \sqrt{c} around the origin. [By the sphere we mean the surface. We generally use the word ball for the inside of the sphere, or the surface together with the inside.] When $c = 0$ there is only the one point $(x, y, z) = (0, 0, 0)$ (the origin) on the 'level surface' $f(x, y, z) = c$. We see then that the level set is not always a 'surface' but it typically is a surface if there are any points at all on it. In this example, when $c < 0$ there are no points satisfying $f(x, y, z) = c$.

2.21 Derivatives of functions of 3 variables. Quite a number of the things we said about derivatives of functions of two variables will also work (more or less in the same way) for functions of 3 variables. There are more variables and so more things to take care about, but most of the ideas are quite similar.

Partial derivatives. The partial derivative of $f(x, y, z)$ with respect to x is denoted $\frac{\partial f}{\partial x}$ — it is computed by differentiating $f(x, y, z)$ with respect to x while treating y and z (the other variables) as though they were constant. We could define it more formally as

$$\frac{\partial f}{\partial x} \Big|_{(x_0, y_0, z_0)} = \frac{d}{dx} \Big|_{x=x_0} f(x, y_0, z_0) = \frac{d}{dt} \Big|_{t=0} f(x_0 + t, y_0, z_0).$$

The other partial derivatives $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ work in a similar way.

There is some merit in the alternative notations for partial derivatives such as $f_x(x_0, y_0, z_0)$ for the value of $\partial f / \partial x$ at the point (x_0, y_0, z_0) . We will generally try to stick to the $\partial f / \partial x$ notation even though it is maybe more clumsy at times.

Directional derivatives are again similar to the two variable case. For a unit vector $\mathbf{u} \in \mathbb{R}^3$ the directional derivative of $f(x, y, z)$ at (x_0, y_0, z_0) in the direction \mathbf{u} is

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \frac{d}{dt} \Big|_{t=0} f((x_0, y_0, z_0) + t\mathbf{u})$$

and we compute it using partial derivatives

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0, z_0)} \right) u_1 + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0, z_0)} \right) u_2 + \left(\frac{\partial f}{\partial z} \Big|_{(x_0, y_0, z_0)} \right) u_3$$

(where u_1, u_2 and u_3 are the components of the unit vector $\mathbf{u} = (u_1, u_2, u_3) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$).

The justification of this formula for the directional derivative depends on the validity of the *linear approximation formula*

$$\begin{aligned} f(x, y, z) &\cong f(x_0, y_0, z_0) + \\ &\quad \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0, z_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0, z_0)} \right) (y - y_0) + \left(\frac{\partial f}{\partial z} \Big|_{(x_0, y_0, z_0)} \right) (z - z_0) \end{aligned}$$

To be more accurate, what we need is that $f(x, y, z)$ should have a *derivative* at the point (x_0, y_0, z_0) , which means that there should be constants a_1, a_2 and a_3 so that

$$\lim_{(x, y, z) \rightarrow (x_0, y_0, z_0)} \frac{f(x, y, z) - (f(x_0, y_0, z_0) + a_1(x - x_0) + a_2(y - y_0) + a_3(z - z_0))}{\text{dist}((x, y, z), (x_0, y_0, z_0))} = 0$$

As in the two variable case we will not go into the definition of a limit in space, but roughly the limit above being 0 means that if (x, y, z) is close enough to (x_0, y_0, z_0) but $(x, y, z) \neq (x_0, y_0, z_0)$, then the fraction above is small.

If the function has a derivative, then the numbers a_1, a_2 and a_3 have to be equal to the partial derivatives $\partial f/\partial x, \partial f/\partial y$ and $\partial f/\partial z$ evaluated at (x_0, y_0, z_0) . Moreover the formula in the box above for directional derivatives will be valid.

How do we know a function has a derivative? Well, as in the 2 variable case there is a theorem that guarantees that most of the functions we can easily write down have derivatives (at least at most points). The theorem says that if the 3 partial derivatives exist and are *continuous functions* at (x_0, y_0, z_0) , then the function has a derivative. We won't go into continuous functions much, but a function $g(x, y, z)$ is called continuous at a point (x_0, y_0, z_0) of its domain if

$$\lim_{(x, y, z) \rightarrow (x_0, y_0, z_0)} g(x, y, z) = g(x_0, y_0, z_0).$$

The function is called continuous (with no mention of a specific point) if it is continuous at each point of its domain. There are theorems about limits of products, quotients and such combinations of functions that allow one to be sure that most formulae define continuous functions (apart perhaps from problem points).

Then there is the *gradient vector*

$$\nabla f \Big|_{(x_0, y_0, z_0)} = \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0, z_0)}, \frac{\partial f}{\partial y} \Big|_{(x_0, y_0, z_0)}, \frac{\partial f}{\partial z} \Big|_{(x_0, y_0, z_0)} \right)$$

or

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

The formula in the box above for the directional derivative can be restated with the gradient vector

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = (\nabla f|_{(x_0, y_0, z_0)}) \cdot \mathbf{u}$$

So, as before in 2 variables, we can say that the directional derivative is

$$D_{\mathbf{u}}f(x_0, y_0, z_0) = \|\nabla f|_{(x_0, y_0, z_0)}\| \cos \theta$$

where θ is the angle between the gradient vector and the direction (unit vector) \mathbf{u} . From this we can say that the directional derivative is largest when \mathbf{u} is the unit vector in the direction of the gradient, that is when

$$\mathbf{u} = \frac{1}{\|\nabla f|_{(x_0, y_0, z_0)}\|} \nabla f|_{(x_0, y_0, z_0)}.$$

And the largest possible value of the directional derivative is the magnitude of the gradient vector. The smallest possible directional derivative occurs when \mathbf{u} is in the exact opposite direction to the gradient vector ($\theta = \pi$ and $\cos \theta = -1$) and this smallest directional derivative is minus the magnitude of the gradient vector.

Another way to put that is that the gradient vector is a vector in the direction of steepest increase of the function and the length of the gradient vector is the rate of increase in that direction.

An application of this that arises fairly frequently is when the function represents potential energy. If a charged particle is subject to an electric field it would have a potential energy $V(x, y, z)$ that depends on its position in the field. If left unchecked by any other influence, the particle would tend to go in the direction of steepest decrease of potential energy. That means that the force acting on the particle would be $\mathbf{F} = -\nabla V$ (and would depend on position (x, y, z)). The same idea would apply if the particle has some mass and is in a gravitational field (caused by the earth, the moon or some other body). The gravitational force acting on it would be minus the gradient (thus in the direction of free fall). If a gas like the atmosphere had a variable pressure then the gas will tend to fill in the low pressure areas. So it will be influenced to go towards the low pressure areas in the direction of fastest decrease of pressure.

A familiar example of this is evident in the weather charts we see every day. The curves of constant pressure (isobars) are shown, but the wind will be almost perpendicular to those lines. It will maybe not quite be perpendicular as there is also an influence of other forces like coreolis forces arising from the rotation of the earth. But this is a seemingly two dimensional example because the weather charts are two dimensional. In fact the atmosphere itself is three dimensional and the air can move vertically as well as horizontally (though this is more likely to be noticed in a plane at higher altitudes than the ground).

An important thing to mention is the property that the gradient is perpendicular to the level surface. This fits with the fact that the directional derivative is zero in the directions perpendicular to the level surface.

We will go on to give some examples of this, but we won't really explain it further. A reasonably good explanation could be given by considering curves $\mathbf{p}(t) = (x(t), y(t), z(t))$ that lie entirely in a level surface $f(x, y, z) = c$ and satisfy $\mathbf{p}(0) = (x_0, y_0, z_0)$. Then differentiate the equation

$$f(x(t), y(t), z(t)) = c$$

with respect to t using the chain rule. This is similar to an explanation we outlined in 2.17 above for level curves of functions of two variables. A difference is that when we are dealing with a level surface there are many possible curves on the surface that pass through the point (x_0, y_0, z_0) and we should look at all of them to see all possible tangent directions. (For each direction in the tangent plane to the surface you can expect to find a curve in the surface that is tangent to that direction as it passes (x_0, y_0, z_0) .)

We did not say yet what the chain rule is in this case, but it is like the two variable version with one extra term:

$$\frac{d}{dt} f(x(t), y(t), z(t)) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt}$$

2.22 Examples. (i) Find the equation of the tangent plane to the surface

$$x^4 + y^2 + z^6 = 26$$

at the point $(2, 3, 1)$.

We can see that we are dealing with a level surface to the function $f(x, y, z) = x^4 + y^2 + z^6$, the level surface $f(x, y, z) = c$ where $c = 26$. According to what we said above the normal vector to the level surface is $\nabla f|_{(2,3,1)}$ and so we can set about computing that gradient vector.

$$\begin{aligned} \frac{\partial f}{\partial x} &= 4x^3 \\ \frac{\partial f}{\partial y} &= 2y \\ \frac{\partial f}{\partial z} &= 6z^5 \end{aligned}$$

Evaluating these at the point $(2, 3, 1)$ we get our normal vector

$$\nabla f|_{(2,3,1)} = (32, 6, 6)$$

So the equation of the plane through $(2, 3, 1)$ perpendicular to that is

$$32(x - 2) + 6(y - 3) + 6(z - 1) = 0$$

or $32x + 6y + 6z = 88$.

- (ii) We can treat a graph $z = f(x, y)$ of a function $f(x, y)$ of 2 variables as a level surface $g(x, y, z) = 0$ of the function $g(x, y, z) = z - f(x, y)$ of 3 variables.

This allows us to calculate the tangent plane to the graph in a different way to the way we did previously. Say we want the tangent plane at the point in the graph where $(x, y) = (x_0, y_0)$. We will need to notice before we finish that the z coordinate at that point on the graph is $z_0 = f(x_0, y_0)$. We have

$$\nabla g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z} \right) = \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1 \right)$$

and we should evaluate this at (x_0, y_0, z_0) to get the normal vector to the tangent plane. This gives the equation to the tangent plane to the graph is

$$\left(-\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(-\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0) + (z - z_0) = 0.$$

We can rearrange that as

$$z = z_0 + \left(\frac{\partial f}{\partial x} \Big|_{(x_0, y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \Big|_{(x_0, y_0)} \right) (y - y_0)$$

and this should look familiar.

2.23 Higher order partial derivatives. Now that we know about partial derivatives we can observe that it is possible to take partial derivatives of partial derivatives. (This is like taking second derivatives of functions of one variable.) For example, say $f(x, y, z) = x^2y + yz + 5xyz$ We can compute

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2xy + 5yz \\ \frac{\partial}{\partial z} \left(\frac{\partial f}{\partial x} \right) &= 5y \\ \frac{\partial f}{\partial z} &= 5xy \\ \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial z} \right) &= 5y \end{aligned}$$

We use the notation

$$\frac{\partial^2 f}{\partial x \partial z}$$

for

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial z} \right)$$

(or sometimes the notation f_{zx}). We can do something similar for third order partials (partials of second order partials) like

$$\frac{\partial^3 f}{\partial y \partial x \partial z} = \frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial x \partial z} \right).$$

There is really just one important thing about higher order partials, apart from what they are. That is that the order (usually) does not matter. So for instance

$$\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x}$$

You will see that in the example where we computed them we did get the same answer, and there is a theorem that says we will usually get the same answer independent of the order we take the partials. Of course it will matter which partials you take and how many times you take the partial with respect to any give variable.

To state the theorem a little carefully (and explain what the word “usually” means), the theorem says that if all the second order partial derivatives of (say) $f(x, y, z)$ are continuous functions, then the order does not matter. So if all of these are continuous

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right), \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right), \frac{\partial}{\partial z} \left(\frac{\partial f}{\partial z} \right), \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right), \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right), \dots$$

then the ‘mixed partials’ don’t depend on the order:

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right), \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial z} \right) = \frac{\partial}{\partial z} \left(\frac{\partial f}{\partial x} \right), \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial z} \right) = \frac{\partial}{\partial z} \left(\frac{\partial f}{\partial y} \right)$$

Typically we can be confident that the theorem about mixed partials will apply when we are dealing with functions given by a formula. The second order partial derivatives will be easy enough to work out using the various rules for differentiation, but even without working them out we know we will end up with a formula for the partials and the second order partials. As most formulae give continuous functions, we should be able to apply the theorem. (The only catch might be cases where there are some bad points for the formulae we get for the second order partial derivatives, bad points that were not really bad for the function we began with.)

We use the notation

$$\frac{\partial^2 f}{\partial y^2}$$

for what we get by taking the partial derivative with respect to y two times, and similarly for the other variables.

2.24 Maxima and minima (general remarks). There are quite a few practical reasons to want to find the maximum or minimum value of something. For example to get the most power out of some machine or do build something at the least possible cost.

Typically there are parameters to be adjusted, changed or designed into things, and the end result may depend on the choices of the parameters in some rather complicated way. But, from a mathematical points of view, we can assume that the way the end result depends on the parameters has been figured out somehow and we are then presented with a problem of finding the largest (or smallest) value of some function of the parameters. Typically there will be some restriction on the feasible choices of the parameters, but as long as there is any scope to vary the parameters we have something to do to find the optimum choice of the parameters.

Now typically there may be very many variables. For example an engine management system would have various things it could control like the air/fuel ratio, the ignition timing and the optimum might depend on the ambient conditions such as the load, engine revs, engine oil temperature, outside temperature and humidity, speed, acceleration and other things. An aeroplane has many more things to manage (flaps, rudder, thrust on several engines, climate control in the cabin and so on). A manufacturing process may have many different choices.

These real world problems are potentially very complicated and would require a lot of background. Just to illustrate the ideas that would be needed to solve such a problem we examine the case where there are either 2 or three parameters. So we consider the question of finding the largest value of $f(x, y)$ for (x, y) in some subset of \mathbb{R}^2 . (Or alternatively to find the largest value of $f(x, y, z)$ for (x, y, z) in some subset of \mathbb{R}^3 .)

Last year you found out about how calculus helps to deal with similar problems where we have only one variable, that is finding the largest value of a function $f(x)$. It may then not be surprising that calculus of several variables (partial derivatives) will be relevant when we have more variables.

To begin with we state a theorem that says that some such problems have a solution. Perhaps it should be obvious in a practical problem that there is bound to be an optimum. At least it is obvious if the problem is sensible. Mathematically we can have a theoretical problem that may or may not have a solution and the theorem can help then. Also, even if our problem is supposed to come from a practical problem where we are sure there should be a solution, having a mathematical reason that there is an optimum can provide a check that our mathematical problem is a realistic reflection of the true problem.

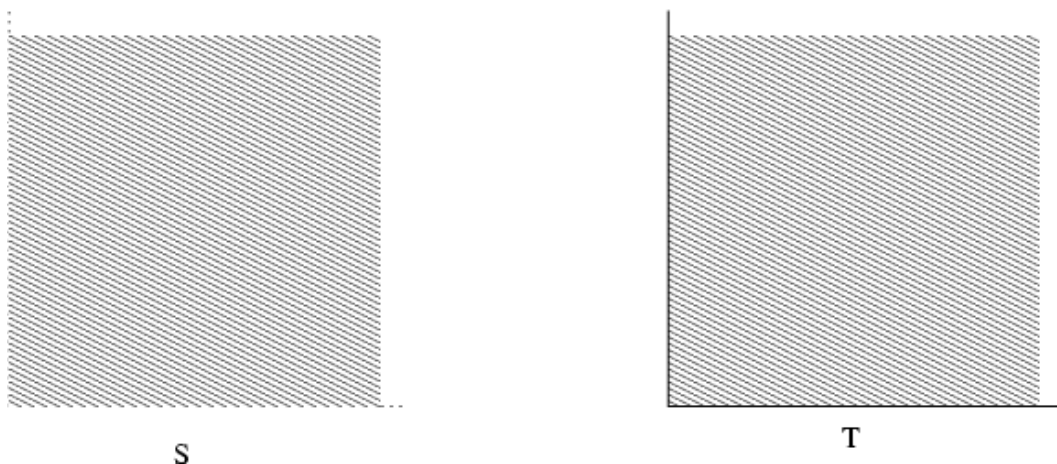
To state the theorem we need to talk about special properties the domain of our function may have called *closed* and *bounded*. In language that maybe seems more precise than it is, a set in \mathbb{R}^2 or \mathbb{R}^3 is called a closed set if it includes all of its ‘boundary points’ or edge points. We could help explain this with some fairly typical examples that are too complicated. The first quadrant

$$S = \{(x, y) \in \mathbb{R}^2 : x > 0 \text{ and } y > 0\}$$

without the bounding axes included is **not** closed, but if we include the edges we do get a closed set. So

$$T = \{(x, y) \in \mathbb{R}^2 : x \geq 0 \text{ and } y \geq 0\}$$

is closed. Here are some pictures (where the axes are supposed to be dotted lines to indicate they are not included on the left set, while the axes are solid on the right to indicate that they are included).



Another example is an disc without its circumference included (which is **not closed**)

$$S = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 2)^2 < 5\}$$

and the version including the circle, which we would call the ‘closed disc’ and is a closed set:

$$T = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 2)^2 \leq 5\}$$

(These discs have centre $(1, 2)$ and radius 5. You might like to sketch them using a dotted circle with the inside shaded to show S and a solid circle with the inside shaded to draw T .)

In general then a set in the plane \mathbb{R}^2 is closed if it includes *all* its boundary curve. We can also make similar examples in space such as the positive ‘orthant’

$$T = \{(x, y, z) \in \mathbb{R}^3 : x \geq 0, y \geq 0 \text{ and } z \geq 0\}.$$

As we are including the bounding quarter planes in this set, T is closed. We could also look at closed balls (the inside of a sphere plus its bounding sphere) as examples of closed sets.

Next we explain what ‘bounded sets’ are. A set $S \subset \mathbb{R}^2$ is called bounded if it does not stretch infinitely far away from the origin. To make it more precise we say that S is *bounded* if there is a finite radius $R > 0$ so that S is contained in the disc of radius R around the origin.

A subset S of \mathbb{R}^3 is called bounded if there is a ball of finite radius that contains S .

So the quadrants are not bounded, but the discs are.

The promised theorem that guarantees the existence of largest and smallest values is this:

Theorem. If S is a closed and bounded set in \mathbb{R}^2 or \mathbb{R}^3 and $f: S \rightarrow \mathbb{R}$ is a *continuous* function, then f has a largest and a smallest value in S .

To spell it out more carefully, there are points $\mathbf{p}_m \in S$ and $\mathbf{p}_M \in S$ so that

$$f(\mathbf{p}_m) \leq f(\mathbf{x}) \leq f(\mathbf{p}_M) \text{ holds for all } \mathbf{x} \in S.$$

The point $\mathbf{x} = \mathbf{p}_M$ is then a point where the value of f is greatest. Or to be even more careful, there is no other point in S where f has a strictly larger value.

This theorem may not seem like much, but you can think of it like this. Finding the largest value of a function can be like finding a needle in a haystack in the sense that it can be very difficult. But knowing that there is definitely a needle to be found in the haystack is a help. If you just search furiously till you get tired and find nothing, you don't know whether there is any needle or not.

Actually, in our case we can use the information that there is a maximum (or minimum), together with a method of narrowing down the possible points where the maximum could be, to provide a way to find the maximum. This is where calculus comes in.

2.25 Critical Points. For a function $f(x)$ of one variable, a critical point is a point where $f'(x) = 0$. For a function $f(x, y)$ of two variables a *critical point* is a point (x_0, y_0) where both partial derivatives are zero:

$$\frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0.$$

For a function $f(x, y, z)$ of three variables a *critical point* is a point (x_0, y_0, z_0) where all three partial derivatives are zero:

$$\frac{\partial f}{\partial x} = 0, \frac{\partial f}{\partial y} = 0 \text{ and } \frac{\partial f}{\partial z} = 0.$$

The relationship of critical points to finding maximum points (or minimum points) is that max and min points have to be critical points **if they are not at the boundary** of the domain. We call points *interior points* of a set if are not at the boundary of the set. (Expressing it more carefully a point $(x_0, y_0) \in S \subseteq \mathbb{R}^2$ is called an interior point of S if all (x, y) within some positive distance of (x_0, y_0) are in S .)

The result is that if a maximum occurs at an interior point of the domain (and the function in question has partial derivatives) then the maximum occurs at a critical point. [This result is actually quite easy to prove based on the analogous result in one variable. When we take partial derivatives we are differentiating only in one variable at a time.]

Since the number of equations to be satisfied (simultaneously) at a critical point is the same as the number of variables (2 equations for the case $f(x, y)$, and 3 equations for the case of $f(x, y, z)$) typically the number of critical points will be quite small. Usually a finite number anyhow. So knowing that the maximum is at a critical points reduces the number of points to be looked at in a dramatic way.

There is a problem, though. We saw just above that a continuous function on a closed and bounded set will have a largest value but there is a problem. The largest value may not be at an interior point — it could also be on the boundary. What we have so far is a way of narrowing down the location of the maximum if it is in the interior, not on the boundary. We'll need a similar method to deal with the boundary case. That will come a little later and is called Lagrange multipliers.

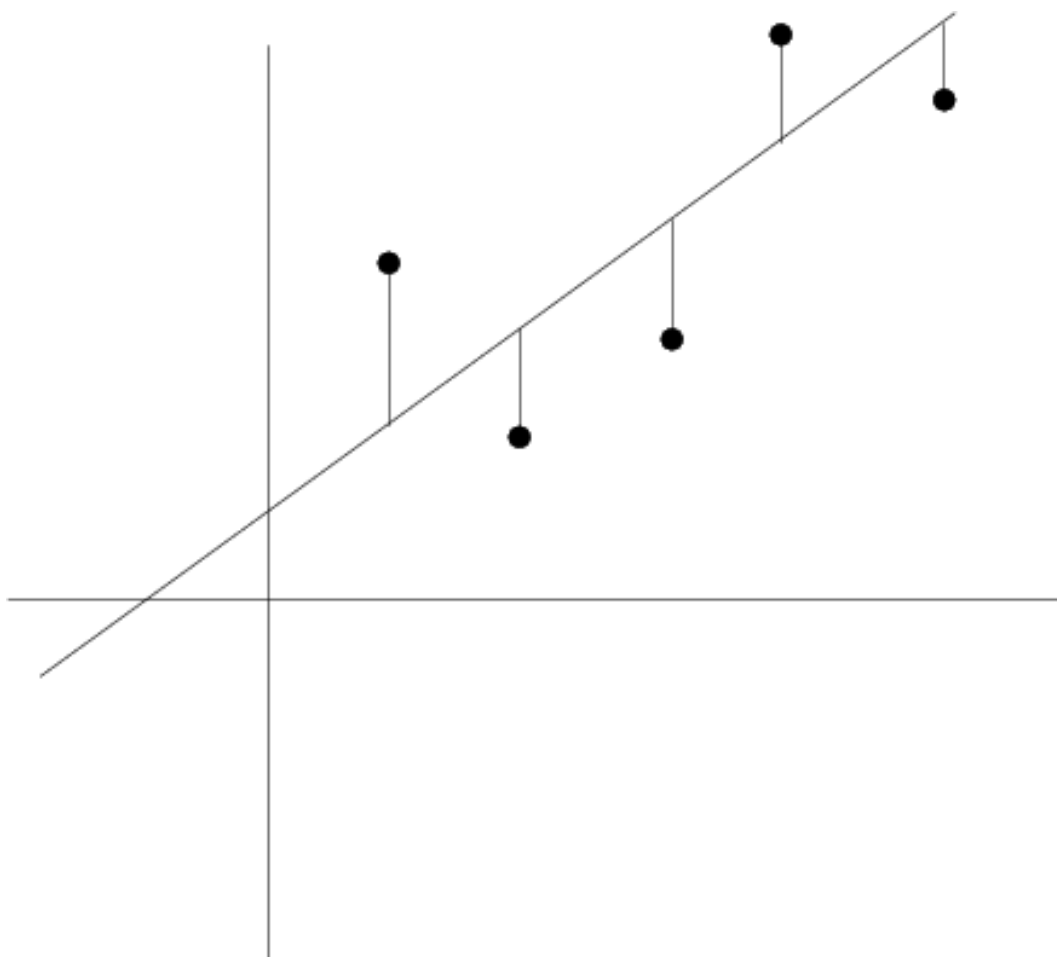
First we will look at an example where there is no boundary

2.26 Example. (Least square fit example.) Assume an experiment that is supposed to produce points on a straight line $y = mx + c$ produces instead a number of data points

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

In the real world, except by the purest fluke, if there are more than 2 points, there will not be any one line that goes through all the data points. The question is to find the line that ‘best fits’ the data. One interpretation of ‘best’ is the line that minimizes the sum of the squares of the vertical discrepancies of the points from the line. So we choose the slope m and the y -intercept c so as to minimize

$$f(m, c) = \sum_{i=1}^n (y_i - (mx_i + c))^2$$



(The picture is supposed to illustrate what we are trying to do. The blobs are the data points and we are trying to choose the line to minimise the sum of the squares of the vertical distances from the points to the line.)

As far as we know now, the right choice of m and c could be any numbers. So we cannot use the theorem guaranteeing that there is an answer to our problem. (The set of possible points (m, c) appears to be all of \mathbb{R}^2 and this is not a bounded set of possibilities.) So we will instead take the view that this is a sensible problem and that there is bound to be an answer.

If this is true, then the minimum will certainly be at an interior point of the allowable (m, c) . Hence it will be at a critical point, a point where both partial derivatives $\partial f / \partial m$ and $\partial f / \partial c$ are

0.²

We now calculate the partial derivatives and set them equal to 0.

$$\begin{aligned}
 \frac{\partial f}{\partial m} &= \sum_{i=1}^n 2(y_i - (mx_i + c))(-x_i) \\
 &= 2 \sum_{i=1}^n (-x_i y_i + mx_i^2 + cx_i) \\
 \frac{\partial f}{\partial c} &= \sum_{i=1}^n 2(y_i - (mx_i + c))(-1) \\
 &= 2 \sum_{i=1}^n (-y_i + mx_i + c) \\
 &= 2 \left(\sum_{i=1}^n (-y_i + mx_i) + \sum_{i=1}^n c \right) \\
 &= 2 \left(\sum_{i=1}^n (-y_i + mx_i) + nc \right)
 \end{aligned}$$

From the equation $\frac{\partial f}{\partial c} = 0$ we find

$$c = -\frac{1}{n} \sum_{i=1}^n (-y_i + mx_i)$$

or

$$c = \frac{1}{n} \left(\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i \right)$$

and when we put that into the equation $\frac{\partial f}{\partial m} = 0$ we get an equation for m only:

$$\begin{aligned}
 \sum_{i=1}^n (-x_i y_i + mx_i^2 + cx_i) &= 0 \\
 - \sum_{i=1}^n (x_i y_i) + m \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i &= 0 \\
 - \sum_{i=1}^n (x_i y_i) + m \sum_{i=1}^n x_i^2 + \frac{1}{n} \left(\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i \right) \sum_{i=1}^n x_i &= 0 \\
 - \sum_{i=1}^n (x_i y_i) + m \sum_{i=1}^n x_i^2 + \frac{1}{n} \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_i \right) - \frac{m}{n} \left(\sum_{i=1}^n x_i \right)^2 &= 0
 \end{aligned}$$

²It is perhaps messy that the variables for our problem are called m and c rather than x and y , but on the other hand we are used to using m for a slope and c for a y -intercept.

If we collect all the terms with m to the left and put the remainder to the right we get

$$m \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) = \sum_{i=1}^n (x_i y_i) - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

This leads to the only choice for m as

$$m = \frac{\sum_{i=1}^n (x_i y_i) - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}$$

and we already had c in terms of m and the data points:

$$c = \frac{1}{n} \left(\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i \right)$$

When you apply the method of least squares to find the best fit line to your data points, what you need is these two formulae. The application of our methods from several variables calculus is one way to arrive at these formulae, but there are others. (This method is given in an exercise in Thomas — in section 14.7 of the 11th edition.)

2.27 Lagrange multipliers. We now give the promised method of Lagrange multipliers. It is aimed at finding the maximum or minimum values of a function $f(x, y)$ subject to a constraint $g(x, y) = 0$.

In fact it does not really do that directly. What it does is provide an analogue of a critical point and narrows down the possible max and minimum points to those that solve several equations. Typically these equations have just a few solutions, and so we are left with only a few candidates for the max and min points.

It is not true any longer that the maximum points have to be at a critical point of f . That result relied on being at an interior point of the allowed (x, y) . Here we are restricted to points on the curve $g(x, y) = 0$ and there will be no interior points of the curve in the plane.

What can say instead is that at the maximum (or minimum) point the gradient of f will be a multiple of the gradient of g . We usually call this multiple by the name λ (Greek letter ‘lambda’). The way to say that in equations is

$$\nabla f = \lambda \nabla g.$$

As this is a vector equation

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = \lambda \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \right)$$

it comes down to two scalar equations

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lambda \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} &= \lambda \frac{\partial g}{\partial y} \end{aligned}$$

We now have 3 unknowns x , y and λ but we also have 3 equations that they should satisfy because we have $g(x, y) = 0$ as well as the two equations above. So we typically get just a few solutions.

That was Lagrange multipliers in 2 dimensions. For 3 dimensions we have a similar thing. The maximum value of a function $f(x, y, z)$ of 3 variables subject to a constraint $g(x, y, z) = 0$ will occur at a place where there is a value of λ that satisfies

$$\begin{aligned}\frac{\partial f}{\partial x} &= \lambda \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} &= \lambda \frac{\partial g}{\partial y} \\ \frac{\partial f}{\partial z} &= \lambda \frac{\partial g}{\partial z} \\ g(x, y, z) &= 0\end{aligned}$$

Now we have 4 equations for the 4 unknowns x , y , z and the Lagrange multiplier λ .

2.28 Example. Find the largest and smallest values of $f(x, y) = xy$ on the ellipse

$$\frac{x^2}{3} + \frac{y^2}{5} = 1$$

Technically the ellipse (just the curve without the inside) is a closed set because it includes all of its boundary. (In fact it is its own boundary, which is a bit confusing perhaps.) The ellipse is also a bounded set and the function $f(x, y)$ is continuous (everywhere and so on that set). That means there must be a solution to our problem — both a largest and a smallest value of $f(x, y)$ on the ellipse must exist.

To narrow down where the maximum and minimum occurs, we use Lagrange multipliers. We take $g(x, y) = x^2/3 + y^2/5 - 1$ so that the ellipse is $g(x, y) = 0$. The maximum (and minimum) must occur at points that solve

$$\nabla f = \lambda \nabla g \text{ and } g(x, y) = 0$$

for some λ . We work out what $\nabla f = \lambda \nabla g$ says and we get

$$(y, x) = \lambda(2x/3, 2y/5)$$

So $y = 2\lambda x/3$ and $x = 2\lambda y/5$. If we put the first of these in the second we get

$$x = \frac{4\lambda^2}{15}x$$

or

$$x \left(1 - \frac{4\lambda^2}{15} \right) = 0$$

Thus there are two cases to look at — $x = 0$ or $\lambda^2 = 15/4$ (which means $\lambda = \pm\sqrt{15}/2$).

If $x = 0$, then we must have $g(x, y) = 0$ and that leaves us with $y^2/5 = 1$, or $y = \pm\sqrt{5}$. Then $(y, x) = \lambda(2x/3, 2y/5)$ comes to $(y, 0) = \lambda(0, 2y/3)$ and there is no λ that can solve this. So $x = 0$ is not an option.

If $\lambda = \pm\sqrt{15}/2$, then $y = 2\lambda x/3 = \pm(\sqrt{15}/3)x = \pm(\sqrt{5/3})x$. So then $g(x, y) = 0$ gives

$$\begin{aligned}\frac{x^2}{3} + \frac{5x^2}{15} &= 1 \\ x^2 \left(\frac{1}{3} + \frac{1}{3} \right) &= 1 \\ x^2 &= \frac{3}{2} \\ x &= \pm\sqrt{\frac{3}{2}}\end{aligned}$$

and $y = 2\lambda x/3 = \pm(\sqrt{5/3})\sqrt{3/2} = \pm\sqrt{5/2}$.

We are left with 4 possible points $(\sqrt{3/2}, \sqrt{5/2})$, $(\sqrt{3/2}, -\sqrt{5/2})$, $(-\sqrt{3/2}, \sqrt{5/2})$ and $(-\sqrt{3/2}, -\sqrt{5/2})$. The values of $f(x, y) = xy$ at these 4 points are $\sqrt{15}/2$, $-\sqrt{15}/2$, $-\sqrt{15}/2$ and $\sqrt{15}/2$. Thus the largest possible value has to be $\sqrt{15}/2$ (which occurs at the two points $(\sqrt{3/2}, \sqrt{5/2})$ and $(-\sqrt{3/2}, -\sqrt{5/2})$). The smallest value is $-\sqrt{15}/2$ (and it occurs at the points $(\sqrt{3/2}, -\sqrt{5/2})$ and $(-\sqrt{3/2}, \sqrt{5/2})$).

2.29 Second derivative test. We now turn to the generalisation to functions of two variables of the familiar second derivative test concerning functions of one variable. To get ourselves straight let us recall that one variable test. We've mentioned above that a maximum point of a function $f(x)$ at an interior point c of the domain will be a critical point (that is a point where $f'(c) = 0$). However, you may recall that this is true not just of absolute maximum points but also of *local maximum* points (which are often called 'relative maximum' points. Similarly local minimum points are critical points if they are not at an end point.

Suppose $x = c$ is a critical point of $f(x)$. The second derivative test says

- If $f'(c) > 0$, then $x = c$ is a local minimum point;
- If $f'(c) < 0$, then $x = c$ is a local maximum point;
- But if $f'(c) = 0$, then no conclusion can be reached without further information.

When we go to 2 variables, we still have the idea of *local maximum* and *local minimum points*. If you think of a graph $z = f(x, y)$ as a landscape, then a local maximum point (x_0, y_0) corresponds to the top of a hill on the graph. It also means $f(x_0, y_0) \geq f(x, y)$ holds for (x, y) within some (possibly small) distance of (x_0, y_0) . A local maximum point does not have to correspond to the top of the highest mountain — that would correspond to the overall largest value and we can call this an *absolute maximum* point.

We can also describe local minimum points as corresponding to the bottom of a valley on the graph. What we can say is that both local max points and local min points will be critical points

of $f(x, y)$ if they are at interior points of the domain. So it is not only absolute max and absolute min points, but also local max and local min points that must be critical points. In one variable there are critical points that are neither (local) max nor min points and there are more critical points where the second derivative test does not tell you what is going on.

In two variables, there are not just peaks and valleys on the graph that are critical points, there are also mountain pass points. If you think of a road across a mountain range, it will usually go over a pass (the lowest point where you can cross the mountain range). As you drive across the pass, there will be a highest point you reach on the road, but as you look to the left and right the mountains will rise on either side of you. So if you imagine that the mountain range is stretching from north to south and the road is heading due east, you see that along the easterly road you are at your highest point when you cross the pass, but if you imagine walking in the northerly direction perpendicular to the road you would be at your lowest point as you cross the road. This ensures that both the partial derivatives with respect to the x and y directions are 0. In the x direction (the east direction) we think we are at a local maximum, but in the northerly direction (the y direction) we think we are at a local minimum. If we look all around us we see that we are neither at the top of a hill nor at the bottom of a valley. Nevertheless we are at a critical point.

These critical points are called *saddle points* (though maybe they could be called mountain pass points) because if you think of a horse saddle you sit down into it when you think about the front to back direction (so it looks like a local minimum in that direction), but you are astride the horse with your legs down each side (so looking only from left to right you are at a local maximum).

The second derivative test in two variables applies to critical points (x_0, y_0) where $\partial f/\partial x = \partial f/\partial y = 0$ and one way to explain it is to talk in terms of the matrix of second order partial derivatives

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

evaluated at the critical point. We want to talk about the matrix being strictly positive, or strictly negative and the way we express this is in terms of both eigenvalues being positive or both being negative.

- If both eigenvalues are strictly positive then (x_0, y_0) is a local minimum point;
- If both eigenvalues are strictly negative then (x_0, y_0) is a local maximum point;
- If one eigenvalue is strictly positive and the other is strictly negative then (x_0, y_0) is a saddle point;
- But if either eigenvalue is zero then no conclusion can be reached without further information.

Rather than computing eigenvalues we can describe this using the determinant of the matrix

$$\begin{aligned} \Delta &= \left(\frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \right) \Big|_{(x_0, y_0)} \\ &= f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2 \end{aligned}$$

using the fact that the determinant is the product of the eigenvalues. (We see then that we should have $\Delta > 0$ in the case where both eigenvalues are strictly positive **and** in the case where both eigenvalues are strictly negative. From $\Delta < 0$ we find that the two eigenvalues have opposite signs.

Suppose (x_0, y_0) is a critical point of $f(x, y)$ and Δ is as above.

- If $\Delta > 0$ and $\frac{\partial^2 f}{\partial x^2} \big|_{(x_0, y_0)} > 0$ then (x_0, y_0) is a local minimum point;
- If $\Delta > 0$ and $\frac{\partial^2 f}{\partial x^2} \big|_{(x_0, y_0)} < 0$ then (x_0, y_0) is a local maximum point;
- If $\Delta < 0$ then (x_0, y_0) is a saddle point;
- But if $\Delta = 0$ then no conclusion can be reached without further information.

2.30 Taylor’s theorem. We turn now to Taylor’s theorem in two variables, which can be used to prove that the second derivative test is right.

Recall for a moment what Taylor’s theorem says for functions of one variable. We can begin by saying that it makes precise the linear approximation formula

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0) \quad (\text{for } x \text{ near } x_0).$$

Or at least it provides a way of figuring out how accurate or inaccurate the formula is in a given case (where the function f and the point x_0 are specified). The theorem says

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(c)(x - x_0)^2$$

for some c between x_0 and x (provided $f''(x)$ exists for x between x_0 and x).

In fact this is just a baby version of Taylor’s theorem, dealing with the ‘remainder term’ in linear approximation. There are also versions dealing with quadratic approximation, cubic approximation, and so on. In general

$$\begin{aligned} f(x) = & f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \cdots + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n \\ & + \frac{1}{(n+1)!}f^{(n+1)}(c)(x - x_0)^{n+1} \end{aligned}$$

(where as before c is between x_0 and x).

We will content ourselves with stating the two variable version of the ‘baby theorem’ though there are versions of Taylor’s theorem for functions of any number of variables and up to any degree n . The notation gets a bit complicated and that is why we stick to the simplest case. Recall the linear approximation formula

$$f(x, y) \cong f(x_0, y_0) + \left(\frac{\partial f}{\partial x} \big|_{(x_0, y_0)} \right) (x - x_0) + \left(\frac{\partial f}{\partial y} \big|_{(x_0, y_0)} \right) (y - y_0)$$

which we rewrite using the subscript notation for derivatives:

$$f(x, y) \cong f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

(valid for (x, y) close to (x_0, y_0)).

The quadratic approximation formula would be

$$\begin{aligned} f(x, y) &\cong f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &\quad + \frac{1}{2!} (f_{xx}(x_0, y_0)(x - x_0)^2 + f_{yy}(x_0, y_0)(y - y_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0)) \end{aligned}$$

and Taylor's theorem gives an equation, rather than an approximation, where the second order partials are evaluated at some point \mathbf{c} on the line joining (x_0, y_0) and (x, y) :

$$\begin{aligned} f(x, y) = & f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ & + \frac{1}{2!} (f_{xx}(\mathbf{c})(x - x_0)^2 + f_{yy}(\mathbf{c})(y - y_0)^2 + 2f_{xy}(\mathbf{c})(x - x_0)(y - y_0)) \end{aligned}$$

One way to use this is to estimate the size of the quadratic terms in the above formula (without being able to know \mathbf{c} exactly) so as to get a worse case scenario for the error in the linear approximation formula.

Another way is to prove that the second derivative test works. If (x_0, y_0) is a critical point, then there are no linear terms (or the coefficients $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$). If we can be sure that the quadratic terms definitely add to something positive then we can conclude that $f(x, y) \geq f(x_0, y_0)$. The assumption for the local min case of the second derivative test makes it possible to deduce that the quadratic terms add to something positive if \mathbf{c} is close enough to (x_0, y_0) . This is guaranteed if (x, y) is close to (x_0, y_0) . That's roughly the idea anyhow.