Functions of two variables

Examples: Functions of several variables
$f(x,y) = x^2 + y^2 \;\Rightarrow\; f(1,2) = 5$ etc.
$f(x,y) = xy^2 \mathrm{e}^{x+y}$
$f(x,y,z) = xy \log z$
Ideal gas law: $P = kT/V$.

Dependent and independent variables
In $z = f(x,y)$ we say x, y are independent variables and z is a dependent variable. This indicates that x and y are free to take any values and then z depends on these values. For now it will be clear which are which, later we'll have to take more care.

Graphs
For the function $y = f(x)$: there is one independent variable and one dependent variable, which means we need 2 dimensions for its graph.

Graphing technique:

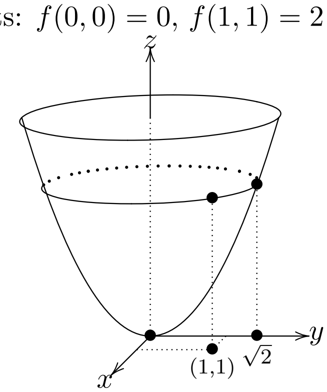
go to x then compute $y = f(x)$ then go up to height y.

For $z = f(x,y)$ we have two independent and one dependent variable, so we need 3 dimensions to graph the function. The technique is the same as before.

Example: Consider $z = f(x,y) = x^2 + y^2$.
To make the graph:

go to (x,y) then compute $z = f(x,y)$ then go up to height z.

We show the plot of three points: $f(0,0) = 0$, $f(1,1) = 2$ and $f(0, \sqrt{2}) = 2$.



The figure above shows more than just the graph of three points. Here are the steps we used to draw the graph. Remember, this is just a sketch, it should suggest the shape of the graph and some of its features.

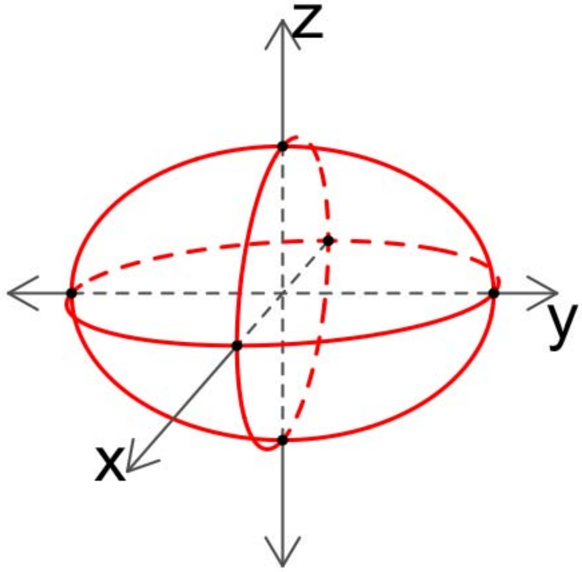
1. First we draw the axes. The z-axis points up, the y-axis is to the right and the x-axis comes out of the page, so it is drawn at the angle shown. This gives a perspective with the eye somewhere in the first octant.

2. The yz-traces are those curves found by setting $x =$ a constant. We start with the trace when $x = 0$. This is an upward pointing parabola in the yz-plane.

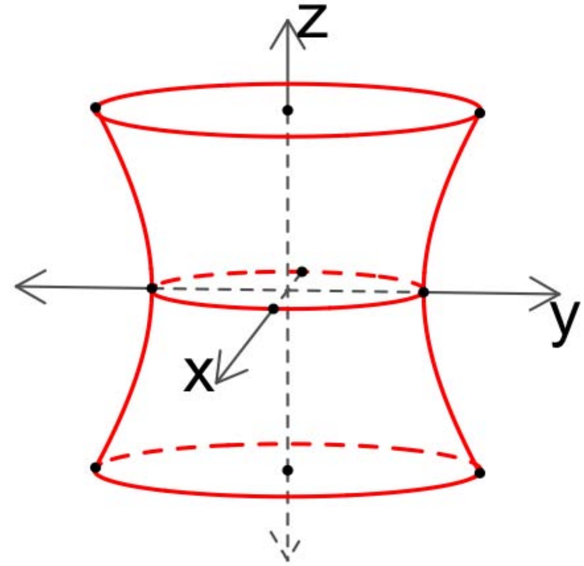
3. Next we sketch the trace with $z = 3$. This is a circle of radius $\sqrt{3}$ at height $z = 3$. Note, the traces where $z =$ constant are generally called *level curves*.

This is enough for this graph. Other graphs take other traces. You should expect to do a certain amount of trial and error before your figure looks right.

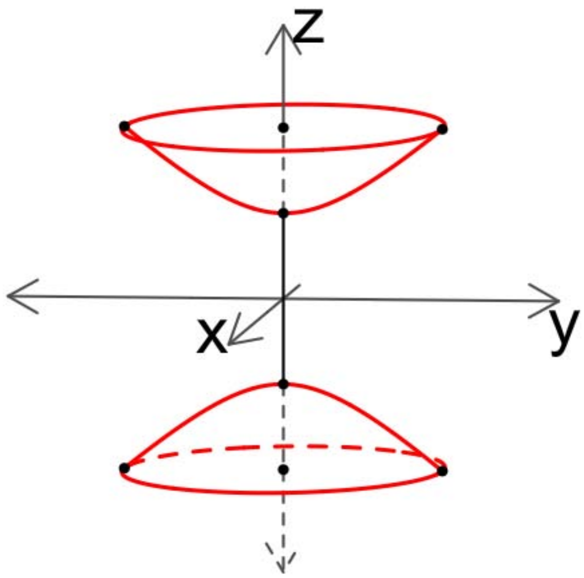
Gallery of graphs



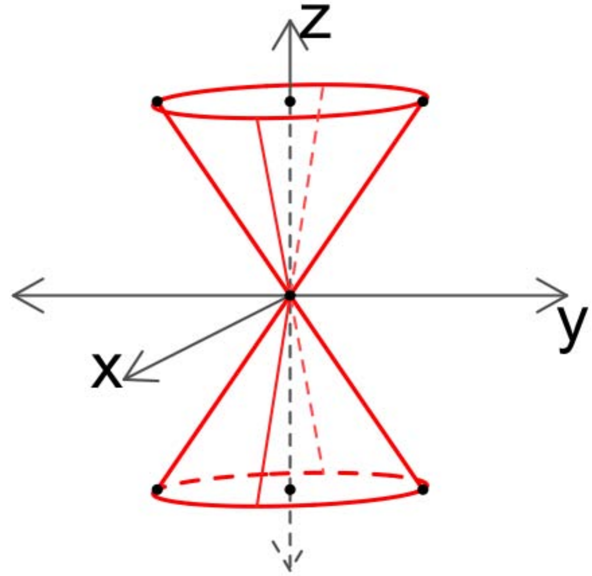
Ellipsoid: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$



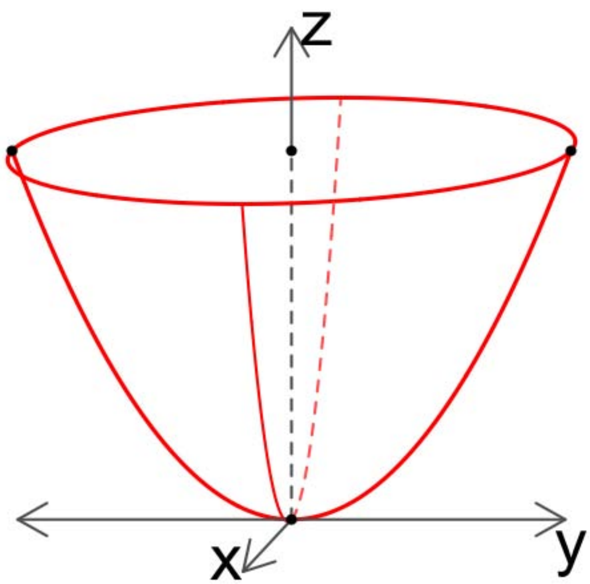
Hyperboloid of one sheet:
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$



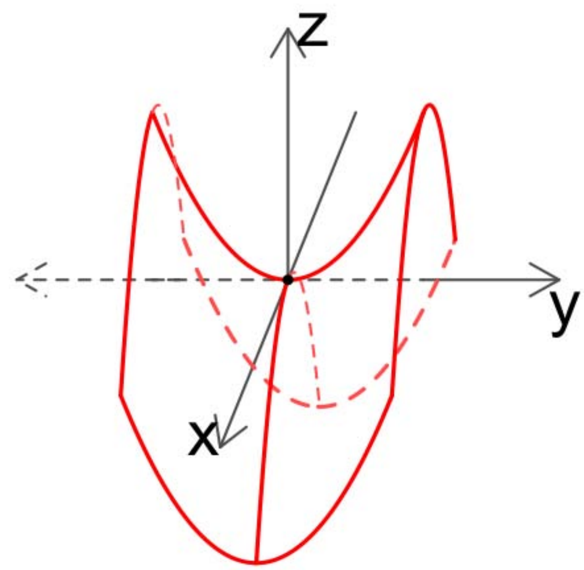
Hyperboloid of two sheets:
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$$



Elliptic cone: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = \dfrac{z^2}{c^2}$



Elliptic paraboloid: $z = ax^2 + by^2$



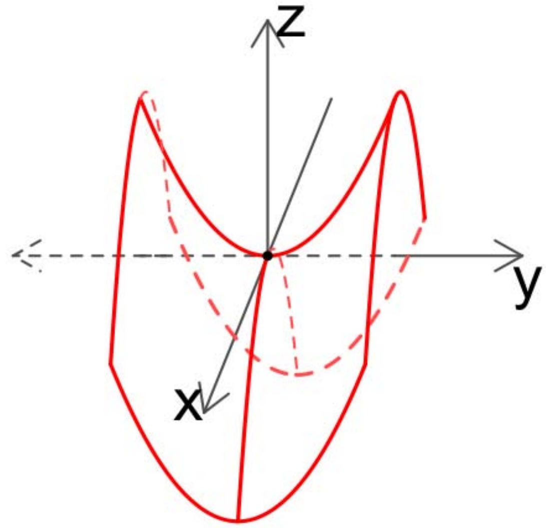
Hyperbolic paraboloid: $z = by^2 - ax^2$

Graphing a function

1. Draw the graph of $z = y^2 - x^2$.

<u>Answer:</u> Here is the graph.
1. First we drew the trace in the yz-plane, which is an upward pointing parabola.
2. Then we drew the xz-traces with $y = 0$, $y = 1$ and $y = -1$.
3. Finally we drew the yz-traces $x = 1$ and $x = -1$.
Lines that are hidden from view are drawn with dashes.

This surface is called a *saddle* and also, a *hyperbolic paraboloid*.



Graphing a function

1. Draw the graph of $z = y^2 - x^2$.

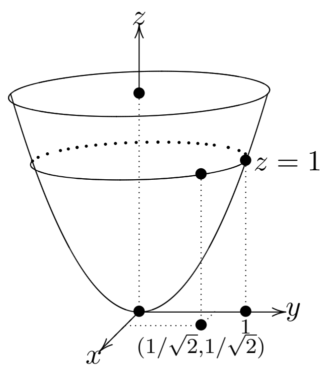
Level Curves and Contour Plots

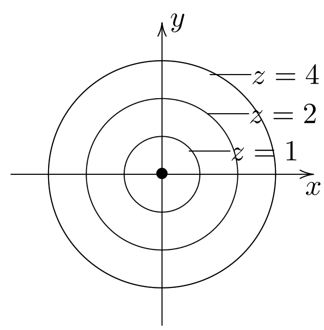
Level curves and *contour plots* are another way of visualizing functions of two variables. If you have seen a topographic map then you have seen a contour plot.

Example: To illustrate this we first draw the graph of $z = x^2 + y^2$. On this graph we draw *contours*, which are curves at a fixed height z = constant.

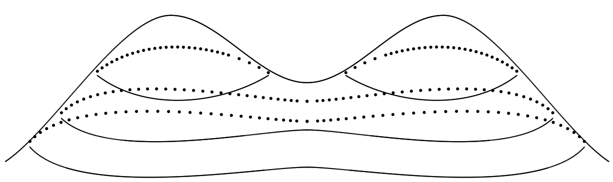
For example the curve at height $z = 1$ is the circle $x^2 + y^2 = 1$. On the graph we have to draw this at the correct height. Another way to show this is to draw the curves in the xy-plane and label them with their z-value. We call these curves *level curves* and the entire plot is called a *contour plot*.

For this example they are shown in the plot on the right. Notice that the 3D graph is simply the level curves 'pulled out' each to its correct height.

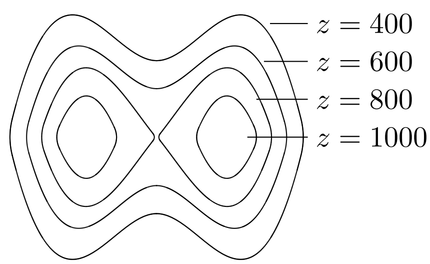




Here is another plot of a 'mountain pass'. Notice that in the contour plot the mountain pass is represented by a level curve that crosses itself. Moving up or down from the cross level curves heights decrease and moving right or left in the other they increase.



Mountain pass

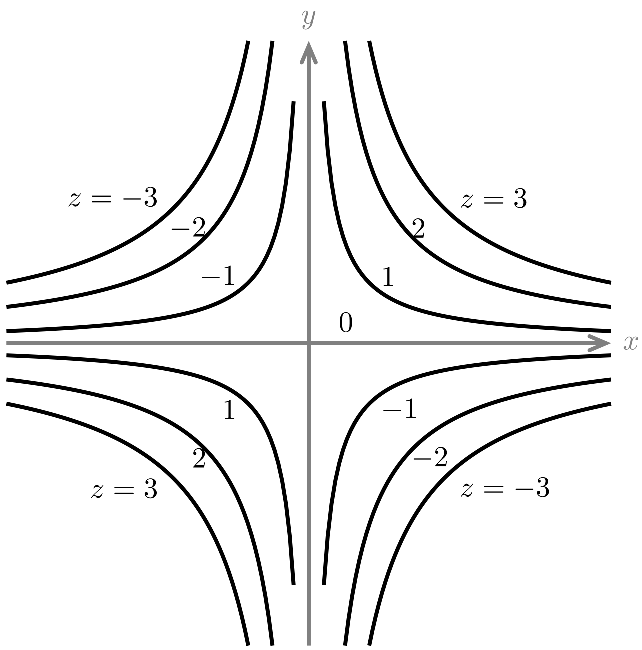


Level curves

Level curves

1. Sketch the level curves for $z = xy$.

Answer: The level curves are hyperbolas $xy =$ constant. We label each curve by the value of the constant.



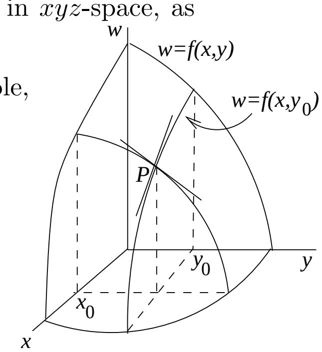
Level curves

1. Sketch the level curves for $z = xy$.

Partial derivatives

Partial derivatives

Let $w = f(x, y)$ be a function of two variables. Its graph is a surface in xyz-space, as pictured.



Fix a value $y = y_0$ and just let x vary. You get a function of *one* variable,

(1) $$w = f(x, y_0), \qquad \text{the \textbf{partial function} for } y = y_0.$$

Its graph is a curve in the vertical plane $y = y_0$, whose slope at the point P where $x = x_0$ is given by the derivative

(2) $$\frac{d}{dx} f(x, y_0)\bigg|_{x_0}, \quad \text{or} \quad \frac{\partial f}{\partial x}\bigg|_{(x_0,y_0)}.$$

We call (2) the **partial derivative** of f with respect to x at the point (x_0, y_0); the right side of (2) is the standard notation for it. The partial derivative is just the ordinary derivative of the partial function — it is calculated by holding one variable fixed and differentiating with respect to the other variable. Other notations for this partial derivative are

$$f_x(x_0, y_0), \qquad \frac{\partial w}{\partial x}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0;$$

the first is convenient for including the specific point; the second is common in science and engineering, where you are just dealing with relations between variables and don't mention the function explicitly; the third and fourth indicate the point by just using a single subscript.

Analogously, fixing $x = x_0$ and letting y vary, we get the partial function $w = f(x_0, y)$, whose graph lies in the vertical plane $x = x_0$, and whose slope at P is the *partial derivative of f with respect to y*; the notations are

$$\frac{\partial f}{\partial y}\bigg|_{(x_0,y_0)}, \qquad f_y(x_0, y_0), \qquad \frac{\partial w}{\partial y}\bigg|_{(x_0,y_0)}, \qquad \left(\frac{\partial f}{\partial y}\right)_0, \qquad \left(\frac{\partial w}{\partial y}\right)_0.$$

The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ depend on (x_0, y_0) and are therefore functions of x and y.

Written as $\partial w/\partial x$, the partial derivative gives the rate of change of w with respect to x alone, at the point (x_0, y_0): it tells how fast w is increasing as x increases, when y is held constant.

For a function of three or more variables, $w = f(x, y, z, \dots)$, we cannot draw graphs any more, but the idea behind partial differentiation remains the same: to define the partial derivative with respect to x, for instance, hold all the other variables constant and take the ordinary derivative with respect to x; the notations are the same as above:

$$\frac{d}{dx} f(x, y_0, z_0, \dots) \;=\; f_x(x_0, y_0, z_0, \dots), \qquad \left(\frac{\partial f}{\partial x}\right)_0, \qquad \left(\frac{\partial w}{\partial x}\right)_0.$$

Partial derivatives

1. Let $f(x,y) = \mathrm{e}^{(x^2+y^2)} + x^2 + y^2 + xy + 2y + 3$.

a) Compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

b) Show the second partials can be computed in any order. That is,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

c) Find $\dfrac{\partial f}{\partial x}(1,3)$.

Answer: a) $\dfrac{\partial f}{\partial x} = 2x\mathrm{e}^{(x^2+y^2)} + 2x + y, \quad \dfrac{\partial f}{\partial y} = 2y\mathrm{e}^{(x^2+y^2)} + 2y + x + 2.$

b) To compute $\dfrac{\partial^2 f}{\partial x \partial y}$ we compute the partial with respect to x of $\dfrac{\partial f}{\partial y}$.

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\left(2y\mathrm{e}^{(x^2+y^2)} + 2y + x + 2\right) = 4xy\mathrm{e}^{(x^2+y^2)} + 1.$$

Likewise

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}\left(2x\mathrm{e}^{(x^2+y^2)} + 2x + y\right) = 4xy\mathrm{e}^{(x^2+y^2)} + 1.$$

We have shown the order of differentiation didn't matter.

c) Evaluating $\dfrac{\partial f}{\partial x}(1,3) = 2\mathrm{e}^{10} + 5$.

Partial derivatives

1. Let $f(x,y) = \mathrm{e}^{(x^2+y^2)} + x^2 + y^2 + xy + 2y + 3$.

a) Compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

b) Show the second partials can be computed in any order. That is,

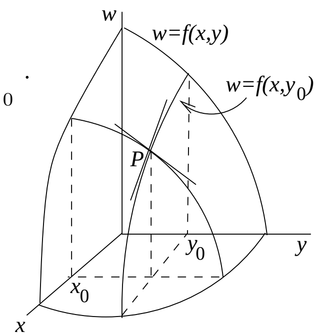
$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

c) Find $\dfrac{\partial f}{\partial x}(1,3)$.

The Tangent Approximation

1. The tangent plane.

For a function of one variable, $w = f(x)$, the tangent line to its graph at a point (x_0, w_0) is the line passing through (x_0, w_0) and having slope $\left(\dfrac{dw}{dx}\right)_0$.



For a function of two variables, $w = f(x, y)$, the natural analogue is the *tangent plane* to the graph, at a point (x_0, y_0, w_0). What's the equation of this tangent plane? Referring to the picture at right (this figure was also used when we introduced partial derivatives), we see that the tangent plane

(i) must pass through (x_0, y_0, w_0), where $w_0 = f(x_0, y_0)$;

(ii) must contain the tangent lines to the graphs of the two partial functions — this will hold if the plane has the same slopes in the $\mathbf{i}$ and $\mathbf{j}$ directions as the surface does.

Using these two conditions, it is easy to find the equation of the tangent plane. The general equation of a plane through (x_0, y_0, w_0) is

$$A(x - x_0) + B(y - y_0) + C(w - w_0) \;=\; 0\ .$$

Assume the plane is not vertical; then $C \neq 0$, so we can divide through by C and solve for $w - w_0$, getting

$$w - w_0 \;=\; a(x - x_0) + b(y - y_0), \qquad a = A/C,\ b = B/C. \tag{3}$$

The plane passes through (x_0, y_0, w_0); what values of the coefficients a and b will make it also tangent to the graph there? We have

$$\begin{aligned}
a \;&=\; \text{slope of plane (3) in the } \mathbf{i}\text{-direction} && \text{(by putting } y = y_0 \text{ in (3));}\\
&=\; \text{slope of graph in the } \mathbf{i}\text{-direction,} && \text{(by (ii) above)}\\
&=\; \left(\frac{\partial w}{\partial x}\right)_0; && \text{(by the definition of partial derivative);} \quad \text{similarly,}\\
b \;&=\; \left(\frac{\partial w}{\partial y}\right)_0.
\end{aligned}$$

Therefore the equation of the **tangent plane** to $w = f(x, y)$ at (x_0, y_0) is

$$w - w_0 \;=\; \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0) \tag{4}$$

2. The approximation formula.

The most important use for the tangent plane is to give an approximation that is the basic formula in the study of functions of several variables — almost everything follows in one way or another from it.

The intuitive idea is that if we stay near (x_0, y_0, w_0), the graph of the tangent plane (4) will be a good approximation to the graph of the function $w = f(x, y)$. Therefore if the point (x, y) is close to (x_0, y_0),

$$(5) \qquad \begin{array}{ccc} f(x,y) & \approx & w_0 + \left(\dfrac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\dfrac{\partial w}{\partial y}\right)_0 (y - y_0) \\ \text{height of graph} & \approx & \text{height of tangent plane} \end{array}$$

The function on the right side of (5) whose graph is the tangent plane is often called the **linearization** of $f(x, y)$ at (x_0, y_0): it is the linear function which gives the best approximation to $f(x, y)$ for values of (x, y) close to (x_0, y_0).

An equivalent form of the approximation (5) is obtained by using Δ notation; if we put

$$\Delta x \;=\; x - x_0, \qquad \Delta y \;=\; y - y_0, \qquad \Delta w \;=\; w - w_0,$$

then (5) becomes

$$(6) \qquad \Delta w \;\approx\; \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad \text{if} \quad \Delta x \approx 0,\; \Delta y \approx 0\ .$$

This formula gives the approximate change in w when we make a small change in x and y. We will use it often.

The analogous approximation formula for a function $w = f(x, y, z)$ of three variables would be

$$(7) \qquad \Delta w \;\approx\; \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y + \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \qquad \text{if} \quad \Delta x, \Delta y \Delta z \approx 0\ .$$

Unfortunately, for functions of three or more variables, we can't use a geometric argument for the approximation formula (7); for this reason, it's best to recast the argument for (6) in a form which doesn't use tangent planes and geometry, and therefore can be generalized to several variables. This is done at the end of this Chapter TA; for now let's just assume the truth of (7) and its higher-dimensional analogues.

Here are two typical examples of the use of the approximation formula. Other examples are in the Exercises. In the rest of your study of partial differentiation, you will see how the approximation formula is used to derive the important theorems and formulas.

Example 1. Give a reasonable square, centered at $(1, 1)$, over which the value of $w = x^3y^4$ will not vary by more than $\pm$ .1 .

Solution. We use (6). We calculate for the two partial derivatives

$$w_x = 3x^2y^4 \qquad\qquad w_y = 4x^3y^3$$

and therefore, evaluating the partials at $(1, 1)$ and using (6), we get

$$\Delta w \;\approx\; 3\Delta x + 4\Delta y.$$

Thus if $|\Delta x| \leq .01$ and $\Delta y| \leq .01$, we should have

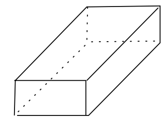
$$|\Delta w| \;\leq\; 3|\Delta x| + 4|\Delta y| \;\leq\; .07,$$

which is within the bounds. So the answer is the square with center at (1,1) given by

$$|x - 1| \leq .01, \qquad |y - 1| \leq .01\ .$$

Example 2. The sides a, b, c of a rectangular box have lengths measured to be respectively 1, 2, and 3. To which of these measurements is the volume V most sensitive?

Solution. $V = abc$, and therefore by the approximation formula (7),



$$\begin{aligned} \Delta V &\approx bc\,\Delta a + ac\,\Delta b + ab\,\Delta c \\ &\approx 6\,\Delta a + 3\,\Delta b + 2\,\Delta c, \qquad \text{at } (1,2,3); \end{aligned}$$

thus it is most sensitive to small changes in side a, since Δa occurs with the largest coefficient. (That is, if one at a time the measurement of each side were changed by say .01, it is the change in a which would produce the biggest change in V, namely .06 .)

The result may seem paradoxical — the value of V is most sensitive to the length of the *shortest* side — but it's actually intuitive, as you can see by thinking about how the box looks.

Sensitivity Principle *The numerical value of $w = f(x, y, \dots)$, calculated at some point $(x_0, y_0, \dots)$, will be most sensitive to small changes in that variable for which the corresponding partial derivative $w_x, w_y, \dots$ has the largest absolute value at the point.*

The Tangent approximation

4. Critique of the approximation formula.

First of all, the approximation formula for functions of two or three variables

$$(6) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad \text{if} \quad \Delta x \approx 0,\ \Delta y \approx 0\ .$$

$$(7) \qquad \Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y + \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \qquad \text{if} \quad \Delta x, \Delta y \Delta z \approx 0\ .$$

is not a precise mathematical statement, since the symbol $\approx$ does not specify exactly how close the quantitites on either side of the formula are to each other. To fix this up, one would have to specify the error in the approximation. (This can be done, but it is not often used.)

A more fundamental objection is that our discussion of approximations was based on the assumption that the tangent plane is a good approximation to the surface at (x_0, y_0, w_0). Is this really so?

Look at it this way. The tangent plane was determined as the plane which has the same slope as the surface in the $\mathbf{i}$ and $\mathbf{j}$ directions. This means the approximation (6) will be good if you move away from (x_0, y_0) in the $\mathbf{i}$ direction (by taking $\Delta y = 0$), or in the $\mathbf{j}$ direction (putting $\Delta x = 0$). But does the tangent plane have the same slope as the surface in all the other directions as well?

Intuitively, we should expect that this will be so if the graph of $f(x, y)$ is a "smooth" surface at (x_0, y_0) — it doesn't have any sharp points, folds, or look peculiar. Here is the mathematical hypothesis which guarantees this.

Smoothness hypothesis. We say $f(x, y)$ is **smooth** at (x_0, y_0) if

(8) $\qquad$ *f_x and f_y are continuous in some rectangle centered at (x_0, y_0).*

If (8) holds, the approximation formula (6) will be valid.

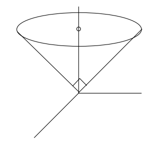
Though pathological examples can be constructed, in general the normal way a function fails to be smooth (and in turn (6) fails to hold) is that one or both partial derivatives fail to exist at (x_0, y_0). This means of course that you won't even be able to write the formula (6), unless you're sleepy. Here is a simple example.

Example 3. Where is $w = \sqrt{x^2 + y^2}$ smooth? Discuss.

Solution. Calculating formally, we get

$$\frac{\partial w}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}}\ , \qquad \frac{\partial w}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}}.$$

These are continuous at all points except $(0, 0)$, where they are undefined. So the function is smooth except at the origin; the approximation formula (6) should be valid everywhere except at the origin.



Indeed, investigating the graph of this function, since $w = \sqrt{x^2 + y^2}$ says that

$$\text{height of graph over } (x, y) \;=\; \text{distance of } (x, y) \text{ from } w\text{-axis},$$

the graph is a right circular cone, with vertex at $(0, 0)$, axis along the w-axis, and vertex angle a right angle. Geometrically the graph has a sharp point at the origin, so there should be no tangent plane there, and no valid approximation formula (6) — there is no linear function which approximates a cone at its vertex.

A non-geometrical argument for the approximation formula

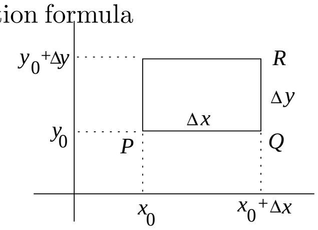
We promised earlier a non-geometrical approach to the approximation formula (6) that would generalize to higher-dimensions, in particular to the 3-variable formula (7). This approach will also show why the hypothesis (8) of *smoothness* is needed. The argument is still imprecise, since it uses the symbol $\approx$, but it can be refined to a proof (which you will find in your book, though it's not easy reading).

It uses the one-variable approximation formula for a differentiable function $w = f(u)$:

$$\Delta w \approx f'(u_0)\Delta u, \qquad \text{if } \Delta u \approx 0\ . \tag{9}$$

We wish to justify — without using reasoning based on 3-space — the approximation formula

$$\Delta w \;\approx\; \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \qquad\qquad \text{if} \quad \Delta x \approx 0,\ \Delta y \approx 0\ . \tag{6}$$



We are trying to calculate the change in w as we go from P to R in the picture, where $P = (x_0, y_0)$, $R = (x_0 + \Delta x, y_0 + \Delta y)$. This change can be thought of as taking place in two steps:

$$\Delta w \;=\; \Delta w_1 \;+\; \Delta w_2, \tag{10}$$

the first being the change in w as you move from P to Q, the second the change as you move from Q to R. Using the one-variable approximation formula (9) :

$$\Delta w_1 \;\approx\; \frac{d}{dx}\, f(x, y_0)\bigg|_{x_0} \cdot \Delta x \;=\; f_x(x_0, y_0)\, \Delta x; \tag{11}$$

similarly,

$$\begin{aligned} \Delta w_2 \;&\approx\; \frac{d}{dy}\, f(x_0 + \Delta x, y)\bigg|_{y_0} \cdot \Delta y \;=\; f_y(x_0 + \Delta x, y_0)\, \Delta y \\ &\approx f_y(x_0, y_0)\, \Delta y, \end{aligned} \tag{12}$$

if we assume that f_y is continuous (i.e., f is smooth), since the difference between the two terms on the right in the last two lines will then be like $\epsilon\, \Delta y$, which is negligible compared with either term itself. Substituting the two approximate values (11) and (12) into (10) gives us the approximation formula (6). □

To make this a proof, the error terms in the approximations have to be analyzed, or more simply, one has to replace the $\approx$ symbol by equalities based on the Mean-Value Theorem of one-variable calculus.

This argument readily generalizes to the higher-dimensional approximation formulas, such as (7); again the essential hypothesis would be smoothness: the three partial derivatives w_x, w_y, w_z should be continuous in a neighborhood of the point (x_0, y_0, z_0).

Tangent approximation

1. Find the equation of the tangent plane to the graph of $z = xy^2$ at the point (1,1,1).

Answer: $\dfrac{\partial z}{\partial x} = y^2$ and $\dfrac{\partial z}{\partial y} = 2xy \ \Rightarrow\ \dfrac{\partial z}{\partial x}(1,1) = 1$ and $\dfrac{\partial z}{\partial y}(1,1) = 1$.

The tangent plane at (1,1,1) is

$$(z-1) = \left.\frac{\partial z}{\partial x}\right|_0 (x-1) + \left.\frac{\partial z}{\partial y}\right|_0 (y-1) = (x-1) + 2(y-1).$$

2. Give the linearization of $f(x,y) = \mathrm{e}^x + x + y$ at (0,0).

Answer: The tangent approximation formula at the point (x_0, y_0, z_0) is

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

(We usually abbreviate this as $\Delta z \approx f_x|_0 \Delta x + f_y|_0 \Delta y$.)

Linearization is just the following form of the tangent approximation formula

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

In our case,

$$f_x(x,y) = \mathrm{e}^x + 1 \ \text{ and } \ f_y(x,y) = 1 \ \Rightarrow\ f(0,0) = 1,\ f_x(0,0) = 2,\ f_y(0,0) = 1$$

Thus, for $(x,y) \approx (0,0)$ we have

$$f(x,y) \approx 1 + 2x + y.$$

Tangent approximation

1. Find the equation of the tangent plane to the graph of $z = xy^2$ at the point (1,1,1).

2. Give the linearization of $f(x, y) = \mathrm{e}^x + x + y$ at (0,0).

Tangent approximation

1. a) Find the equation tangent plane to the graph of $z = x^2 + y^2$ at the point (2,1,5).

b) Give the tangent approximation for z near the point $(x_0, y_0) = (2, 1)$.

Answer: a) $\dfrac{\partial z}{\partial x} = 2x$ and $\dfrac{\partial z}{\partial y} = 2y \Rightarrow \dfrac{\partial z}{\partial x}(2,1) = 4$ and $\dfrac{\partial z}{\partial y}(2,1) = 2$.

The tangent plane at (2,1,5) is

$$(z - 5) = \left.\frac{\partial z}{\partial x}\right|_0 (x - 2) + \left.\frac{\partial z}{\partial y}\right|_0 (y - 1) = 4(x - 2) + 2(y - 1).$$

b) The tangent approximation is the same formula, with the interpretation that for a fixed (x_0, y_0) the value of z on the graph of the function is near that of z on the tangent plane. Thus, for $(x_0, y_0) \approx (2, 1)$ we have

$$\Delta z \approx 4\Delta x + 2\Delta y.$$

Critical Points

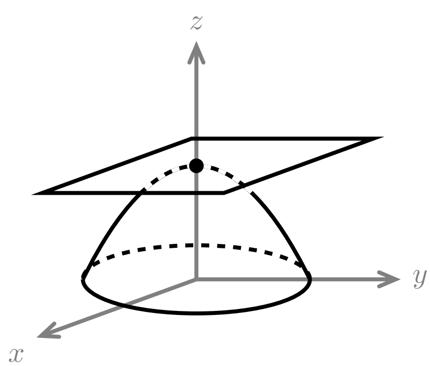
Critical points:
A standard question in calculus, with applications to many fields, is to find the points where a function reaches its relative maxima and minima.

Just as in single variable calculus we will look for maxima and minima (collectively called *extrema*) at points (x_0, y_0) where the first derivatives are 0. Accordingly we define a *critical point* as any point (x_0, y_0) where

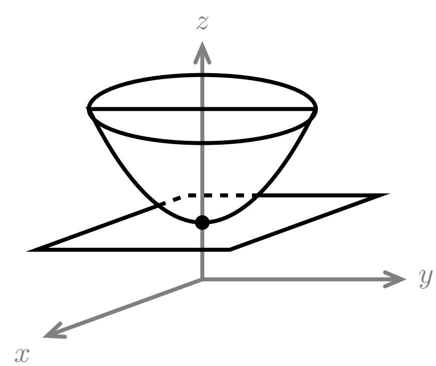
$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \ \text{ and } \ \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Often we will abbreviate this as $f_x = 0$ and $f_y = 0$.

Our first job is to verify that relative maxima and minima occur at critical points. The figures below illustrates that they occur at places where the tangent plane is horizontal.



Max. with horizontal tang. plane



Min. with horizontal tang. plane

Since horizontal planes are of the form $z =$ constant. and the equation of the tangent plane at (x_0, y_0, z_0) is

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

we see it is horizontal when

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

Thus, extrema occur at critical points. But, just as in single variable calculus, not all critical points are extrema.

Example: Find the critical points of $z = x^2 + y^2 + .5$.

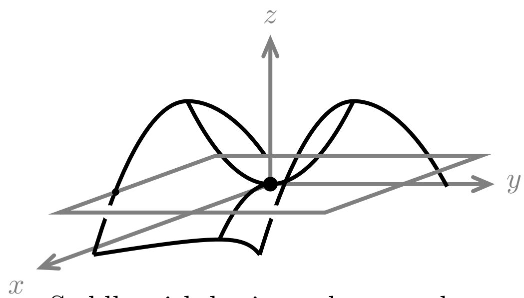
<u>Answer:</u> $\dfrac{\partial z}{\partial x} = 2x$ and $\dfrac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a minimum value. (See the figure above on the right.)

Example: Find the critical points of $z = 1 - x^2 - y^2$.

<u>Answer:</u> $\dfrac{\partial z}{\partial x} = -2x$ and $\dfrac{\partial z}{\partial y} = -2y$. Clearly the only point where both derivatives are 0 is $(0, 0)$. Thus, there is a single critical point at $(0, 0)$. The figure shows it is clearly the point where z reaches a maximum value. (See the figure above on the left.)

Example: Find the critical points of $z = -x^2 + y^2$.

Answer: $\dfrac{\partial z}{\partial x} = -2x$ and $\dfrac{\partial z}{\partial y} = 2y$. Clearly the only point where both derivatives are 0 is $(0,0)$. Thus, there is a single critical point at $(0,0)$. The figure shows it is neither a minimum or a maximum.



Saddle with horizontal tang. plane

Example: Making a box with minimum material.

A box is made of cardboard with double thick sides, a triple thick bottom, single thick front and back and no top. It's volume = 3.

What dimensions use the least amount of cardboard?

Answer: The box shown has dimensions x, y, and z.

The area of one side $= yz$. There are two double thick sides $\Rightarrow$ cardboard used $= 4yz$.

The area of the front (and back) $= xz$. It is single thick $\Rightarrow$ cardboard used $= 2xz$.

The area of the bottom $= xy$. It is triple thick $\Rightarrow$ cardboard used $= 3xy$.

Thus, the total cardboard used is

$$w = 4yz + 2xz + 3xy.$$

The volume $= 3 = xyz \Rightarrow z = \frac{3}{xy}$. Substituting this in the formula for w gives

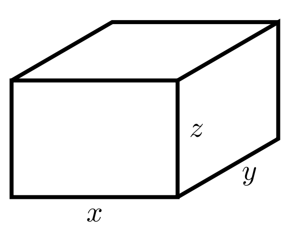
$$w = \frac{12}{x} + \frac{6}{y} + 3xy.$$

We find the critical points of w.

$$w_x = -\frac{12}{x^2} + 3y = 0, \quad w_y = -\frac{6}{y^2} + 3x = 0.$$

The first equation implies $y = \dfrac{4}{x^2}$. Substituting this in the second equation gives $-\dfrac{6}{16}x^4 + 3x = 0$. Thus, $x = 0$ or 2. We reject 0 since then y is undefined. Using $x = 2$ we find $y = 1$. Thus, there there is one critical point at (2,1). and at this point we have $z = 3/2$.

This point gives the box with minimum cardboard used because physically we know it must have a minimum somewhere. Later we will learn to check this with the second derivative test.



Critical points

1. Find all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

Answer: Taking the first partials:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 \text{ and } \frac{\partial z}{\partial y} = 3y^2 - 12.$$

Setting these equal to 0 gives

$$x^5 = -1 \;\Rightarrow\; x = -1 \text{ and } 3y^2 = 12 \;\Rightarrow\; y = \pm 2.$$

The critical points are $(-1, 2)$ and $(-1, -2)$.

Critical points

1. Find all the critical points of

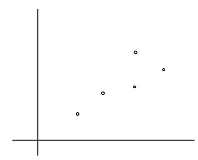
$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

Least Squares Interpolation

1. The least-squares line.

Suppose you have a large number n of experimentally determined points, through which you want to pass a curve. There is a formula (the Lagrange interpolation formula) producing a polynomial curve of degree $n-1$ which goes through the points exactly. But normally one wants to find a simple curve, like a line, parabola, or exponential, which goes approximately through the points, rather than a high-degree polynomial which goes exactly through them. The reason is that the location of the points is to some extent determined by experimental error, so one wants a smooth-looking curve which averages out these errors, not a wiggly polynomial which takes them seriously.

In this section, we consider the most common case — finding a line which goes approximately through a set of data points.



Suppose the data points are

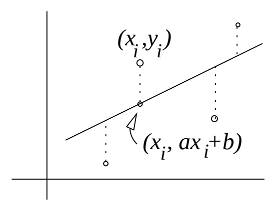
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

and we want to find the line

$$y = ax + b \tag{1}$$

which "best" passes through them. Assuming our errors in measurement are distributed randomly according to the usual bell-shaped curve (the so-called "Gaussian distribution"), it can be shown that the right choice of a and b is the one for which the sum D of the squares of the deviations

$$D = \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2 \tag{2}$$



is a *minimum*. In the formula (2), the quantities in parentheses (shown by dotted lines in the picture) are the **deviations** between the observed values y_i and the ones $ax_i + b$ that would be predicted using the line (1).

The deviations are squared for theoretical reasons connected with the assumed Gaussian error distribution; note however that the effect is to ensure that we sum only positive quantities; this is important, since we do not want deviations of opposite sign to cancel each other out. It also weights more heavily the larger deviations, keeping experimenters honest, since they tend to ignore large deviations ("I had a headache that day").

This prescription for finding the line (1) is called the **method of least squares**, and the resulting line (1) is called the **least-squares** line or the **regression** line.

To calculate the values of a and b which make D a minimum, we see where the two partial derivatives are zero:

$$\begin{aligned} \frac{\partial D}{\partial a} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0 \\ \frac{\partial D}{\partial b} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0\,. \end{aligned} \tag{3}$$

These give us a pair of *linear* equations for determining a and b, as we see by collecting terms and cancelling the 2's:

$$
\begin{aligned}
\left(\sum x_i^2\right) a \;+\; \left(\sum x_i\right) b \;&=\; \sum x_i y_i \\
\left(\sum x_i\right) a \;+\; n\, b \;&=\; \sum y_i \;.
\end{aligned}
\tag{4}
$$

(Notice that it saves a lot of work to differentiate (2) using the chain rule, rather than first expanding out the squares.)

The equations (4) are usually divided by n to make them more expressive:

$$
\begin{aligned}
\bar{s}\, a \;+\; \bar{x}\, b \;&=\; \frac{1}{n}\sum x_i y_i \\
\bar{x}\, a \;+\; b \;&=\; \bar{y}\;,
\end{aligned}
\tag{5}
$$

where $\bar{x}$ and $\bar{y}$ are the average of the x_i and y_i, and $\bar{s} = \sum x_i^2/n$ is the average of the squares.

From this point on use linear algebra to determine a and b. It is a good exercise to see that the equations are always solvable unless all the x_i are the same (in which case the best line is vertical and can't be written in the form (1)).

In practice, least-squares lines are found by pressing a calculator button, or giving a MatLab command. Examples of calculating a least-squares line are in the exercises accompanying the course. Do them from scratch, starting from (2), since the purpose here is to get practice with max-min problems in several variables; don't plug into the equations (5). Remember to differentiate (2) using the chain rule; don't expand out the squares, which leads to messy algebra and highly probable error.

2. Fitting curves by least squares.

If the experimental points seem to follow a curve rather than a line, it might make more sense to try to fit a second-degree polynomial

$$
y = a_0 + a_1 x + a_2 x^2 \tag{6}
$$

to them. If there are only three points, we can do this exactly (by the Lagrange interpolation formula). For more points, however, we once again seek the values of a_0, a_1, a_2 for which the sum of the squares of the deviations

$$
D \;=\; \sum_1^n \big(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\big)^2 \tag{7}
$$

is a minimum. Now there are three unknowns, a_0, a_1, a_2. Calculating (remember to use the chain rule!) the three partial derivatives $\partial D/\partial a_i,\ i = 0, 1, 2$, and setting them equal to zero leads to a square system of three linear equations; the a_i are the three unknowns, and the coefficients depend on the data points (x_i, y_i). They can be solved by finding the inverse matrix, elimination, or using a calculator or MatLab.

If the points seem to lie more and more along a line as $x \to \infty$, but lie on one side of the line for low values of x, it might be reasonable to try a function which has similar behavior, like

$$
y \;=\; a_0 + a_1 x + a_2\,\frac{1}{x} \tag{8}
$$

and again minimize the sum of the squares of the deviations, as in (7). In general, this method of least squares applies to a trial expression of the form

$$y = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_r f_r(x), \tag{9}$$

where the $f_i(x)$ are given functions (usually simple ones like $1, x, x^2, 1/x, e^{kx}$, etc. Such an expression (9) is called a **linear combination** of the functions $f_i(x)$. The method produces a square inhomogeneous system of linear equations in the unknowns $a_0, \ldots, a_r$ which can be solved by finding the inverse matrix to the system, or by elimination.

The method also applies to finding a linear function

$$z = a_1 + a_2 x + a_3 y \tag{10}$$

to fit a set of data points

$$(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n) \ . \tag{11}$$

where there are two independent variables x and y and a dependent variable z (this is the quantity being experimentally measured, for different values of (x, y)). This time after differentiation we get a 3×3 system of linear equations for determining a_1, a_2, a_3 .

The essential point in all this is that the unknown coefficients a_i should occur *linearly* in the trial function. Try fitting a function like ce^{kx} to data points by using least squares, and you'll see the difficulty right away. (Since this is an important problem — fitting an exponential to data points — one of the Exercises explains how to adapt the method to this type of problem.)

Least Squares Interpolation

1. Use the method of least squares to fit a line to the four data points

$$(0,0), \quad (1,2), \quad (2,1), \quad (3,4).$$

Answer: We are looking for the line $y = ax + b$ that best models the data. The deviation of a data point (x_i, y_i) from the model is

$$y_i - (ax_i + b).$$

The sum of the squares of the deviation is

$$\begin{aligned} D &= (0-(a\cdot 0+b))^2 + (2-(a\cdot 1+b))^2 + (1-(a\cdot 2+b))^2 + (4-(a\cdot 3+b))^2 \\ &= b^2 + (2-a-b)^2 + (1-2a-b)^2 + (4-3a-b)^2. \end{aligned}$$

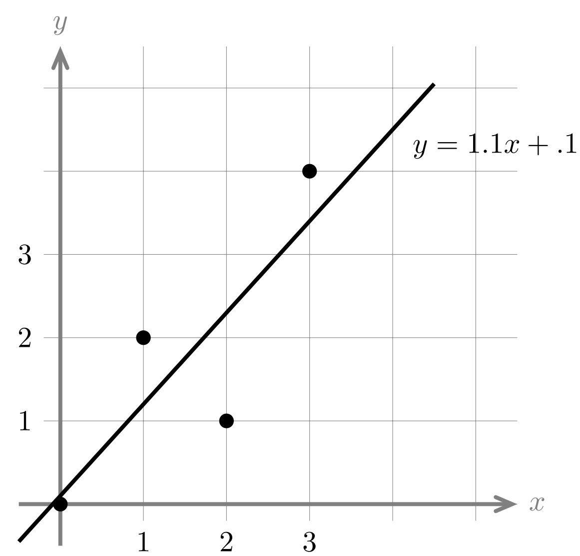
Taking derivatives and setting them to 0 gives

$$\begin{aligned} \tfrac{\partial D}{\partial a} &= -2(2-a-b) - 4(1-2a-b) - 6(4-3a-b) = 0 \;\Rightarrow\; 28a + 12b = 32 \;\Rightarrow\; 14a + 6b = 16 \\ \tfrac{\partial D}{\partial b} &= 2b - 2(2-a-b) - 2(1-2a-b) - 2(4-3a-b) = 0 \;\Rightarrow\; 12a + 8b = 14 \;\Rightarrow\; 6a + 4b = 7. \end{aligned}$$

This linear system of two equations in two unknowns is easy to solve. We get

$$a = \frac{11}{10}, \qquad b = \frac{1}{10}.$$

Here is a plot of the problem.



Least Squares Interpolation

1. Use the method of least squares to fit a line to the four data points

$$(0,0), \quad (1,2), \quad (2,1), \quad (3,4).$$

Least squares interpolation

1. Use the method of least squares to fit a line to the three data points

$$(0,0), \quad (1,2), \quad (2,1).$$

<u>**Answer:**</u> We are looking for the line $y = ax + b$ that best models the data. The deviation of a data point (x_i, y_i) from the model is

$$y_i - (ax_i + b).$$

By best we mean the line that minimizes the sum of the squares of the deviation. That is we want to minimize

$$\begin{aligned} D &= (0 - (a \cdot 0 + b))^2 + (2 - (a \cdot 1 + b))^2 + (1 - (a \cdot 2 + b))^2 \\ &= b^2 + (2 - a - b)^2 + (1 - 2a - b)^2. \end{aligned}$$

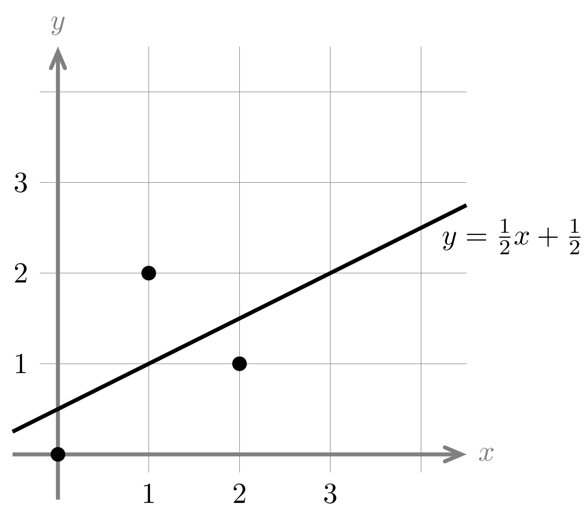
(Remember, the variables whose values are to be found are a and b.) We do not expand out the squares, rather we take the derivatives first. Setting the derivatives equal to 0 gives

$$\begin{aligned} &\tfrac{\partial D}{\partial a} = -2(2 - a - b) - 4(1 - 2a - b) = 0 \;\Rightarrow\; 10a + 6b = 8 \;\Rightarrow\; 5a + 3b = 4 \\ &\tfrac{\partial D}{\partial b} = 2b - 2(2 - a - b) - 2(1 - 2a - b) = 0 \;\Rightarrow\; 6a + 6b = 6 \;\Rightarrow\; 3a + 3b = 3. \end{aligned}$$

This linear system of two equations in two unknowns is easy to solve. We get

$$a = \frac{1}{2}, \qquad b = \frac{1}{2}.$$

Here is a plot of the problem.



Second Derivative Test

1. The Second Derivative Test

We begin by recalling the situation for twice differentiable functions $f(x)$ of one variable. To find their local (or "relative") maxima and minima, we

1. find the critical points, i.e., the solutions of $f'(x) = 0$;
2. apply the second derivative test to each critical point x_0:

$$f''(x_0) > 0 \;\Rightarrow\; x_0 \text{ is a local minimum point;}$$
$$f''(x_0) < 0 \;\Rightarrow\; x_0 \text{ is a local maximum point.}$$

The idea behind it is: at x_0 the slope $f'(x_0) = 0$; if $f''(x_0) > 0$, then $f'(x)$ is strictly increasing for x near x_0, so that the slope is negative to the left of x_0 and positive to the right, which shows that x_0 is a minimum point. The reasoning for the maximum point is similar.

If $f''(x_0) = 0$, the test fails and one has to investigate further, by taking more derivatives, or getting more information about the graph. Besides being a maximum or minimum, such a point could also be a horizontal point of inflection.

The analogous test for maxima and minima of functions of two variables $f(x, y)$ is a little more complicated, since there are several equations to satisfy, several derivatives to be taken into account, and another important geometric possibility for a critical point, namely a **saddle point**. This is a local minimax point; around such a point the graph of $f(x, y)$ looks like the central part of a saddle, or the region around the highest point of a mountain pass. In the neighborhood of a saddle point, the graph of the function lies both above and below its horizontal tangent plane at the point.

The second-derivative test for maxima, minima, and saddle points has two steps.

1. Find the critical points by solving the simultaneous equations $\begin{cases} f_x(x, y) = 0, \\ f_y(x, y) = 0. \end{cases}$

Since a critical point (x_0, y_0) is a solution to both equations, both partial derivatives are zero there, so that the tangent plane to the graph of $f(x, y)$ is horizontal.

2. To test such a point to see if it is a local maximum or minimum point, we calculate the three second derivatives at the point (we use subscript 0 to denote evaluation at (x_0, y_0), so for example $(f)_0 = f(x_0, y_0)$), and denote the values by A, B, and C:

$$A = (f_{xx})_0, \qquad B = (f_{xy})_0 = (f_{yx})_0, \qquad C = (f_{yy})_0, \tag{1}$$

(we are assuming the derivatives exist and are continuous).

Second-derivative test. *Let (x_0, y_0) be a critical point of $f(x, y)$, and A, B, and C be as in (1). Then*

$$AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \;\Rightarrow\; (x_0, y_0) \text{ is a minimum point;}$$
$$AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \;\Rightarrow\; (x_0, y_0) \text{ is a maximum point;}$$
$$AC - B^2 < 0 \;\Rightarrow\; (x_0, y_0) \text{ is a saddle point.}$$

If $AC - B^2 = 0$, the test fails and more investigation is needed.

Note that if $AC - B^2 > 0$, then $AC > 0$, so that A and C must have the same sign.

Example 1. Find the critical points of $w = 12x^2 + y^3 - 12xy$ and determine their type.

Solution. We calculate the partial derivatives easily:

$$
\begin{aligned} w_x &= 24x - 12y \\ w_y &= 3y^2 - 12x \end{aligned} \qquad \begin{aligned} A &= w_{xx} = 24 \\ B &= w_{xy} = -12 \\ C &= w_{yy} = 6y \end{aligned} \tag{2}
$$

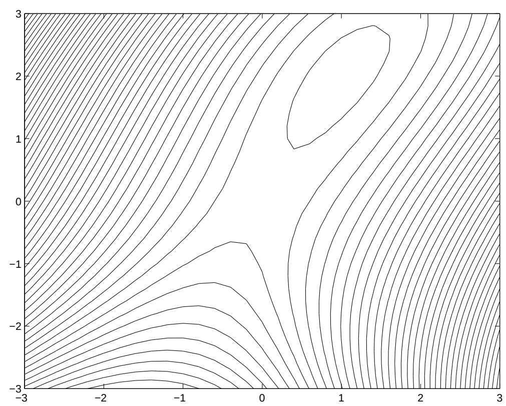
To find the critical points we solve simultaneously the equations $w_x = 0$ and $w_y = 0$; we get

$$
\begin{aligned} w_x &= 0 \\ w_y &= 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} y &= 2x \\ y^2 &= 4x \end{aligned} \quad \Rightarrow \quad 4x^2 = 4x \quad \Rightarrow \quad x = 0, 1 \quad \Rightarrow \quad \begin{aligned} (x,y) &= (0,0) \\ (x,y) &= (1,2) \end{aligned}.
$$

Thus there are two critical points: $(0,0)$ and $(1,2)$. To determine their type, we use the second derivative test: we have $AC - B^2 = 144y - 144$, so that

at $(0,0)$, we have $AC - B^2 = -144$, so it is a saddle point;
at $(1,2)$, we have $AC - B^2 = 144$ and $A > 0$, so it is a a minimum point.

A plot of the level curves is given at the right, which confirms the above. Note that the behavior of the level curves near the origin can be determined by using the approximation $w \approx 12x^2 - 12xy$; this shows the level curves near $(0,0)$ look like those of the function $x(x-y)$: the family of hyperbolas $x(x-y) = c$, with asymptotes given by the degenerate hyperbola $x(x-y) = 0$, i.e., the pair of lines $x = 0$ (the y-axis) and $x - y = 0$ (the diagonal line $y = x$).



2. Justification for the Second-derivative Test.

The test involves the quantity $AC - B^2$. In general, whenever we see the expressions $B^2 - 4AC$ or $B^2 - AC$ or their negatives, it means the quadratic formula is involved, in one of its two forms (the second is often used to get rid of the excess two's):

$$
Ax^2 + Bx + C = 0 \quad \Rightarrow \quad x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \tag{3}
$$

$$
Ax^2 + 2Bx + C = 0 \quad \Rightarrow \quad x = \frac{-B \pm \sqrt{B^2 - AC}}{A} \tag{4}
$$

This is what is happening here. We want to know whether, near a critical point P_0, the graph of our function $w = f(x,y)$ always stays on one side of its horizontal tangent plane (P_0 is then a maximum or minimum point), or whether it lies partly above and partly below the tangent plane (P_0 is then a saddle point). As we will see, this is determined by how the graph of a quadratic function $f(x)$ lies with respect to the x-axis. Here is the basic lemma.

Lemma. For the quadratic function $Ax^2 + 2Bx + C$,

$$
AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 \quad \Rightarrow \quad Ax^2 + 2Bx + C > 0 \quad \text{for all } x; \tag{5}
$$

$$
AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 \quad \Rightarrow \quad Ax^2 + 2Bx + C < 0 \quad \text{for all } x; \tag{6}
$$

$$
AC - B^2 < 0 \quad \Rightarrow \quad \begin{cases} Ax^2 + 2Bx + C > 0, & \text{for some } x; \\ Ax^2 + 2Bx + C < 0, & \text{for some } x. \end{cases} \tag{7}
$$

Proof of the Lemma. To prove (5), we note that the quadratic formula in the form (4) shows that the zeros of $Ax^2 + 2Bx + C$ are imaginary, i.e., it has no real zeros. Therefore its graph must lie entirely on one side of the x-axis; which side can be determined from either A or C, since

$$A > 0 \;\Rightarrow\; \lim_{x\to\infty} Ax^2 + 2Bx + C = \infty; \qquad C > 0 \;\Rightarrow\; Ax^2 + 2Bx + C > 0 \text{ when } x = 0.$$

If $A < 0$ or $C < 0$, the reasoning is analogous and proves (6).

If on the other hand $AC - B^2 < 0$, formula (4) shows the quadratic function has two real roots, so that its parabolic graph crosses the x-axis twice, and hence lies partly above and partly below it. This proves (7). □

Proof of the Second-derivative Test in a special case.

The simplest function is a linear function, $w = w_0 + ax + by$, but it does not in general have maximum or minimum points and its second derivatives are all zero. The simplest functions to have interesting critical points are the quadratic functions, which we write in the form (the 2's will be explained momentarily):

$$w = w_0 + ax + by + \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{8}$$

Such a function has in general a unique critical point, which we will assume is $(0,0)$; this gives the function a special form, which we can determine by evaluating its partial derivatives at $(0,0)$:

$$\begin{array}{ll} (w_x)_0 = a & \quad w_{xx} = A \\ (w_y)_0 = b & \quad w_{xy} = B \\ & \quad w_{yy} = C \end{array} \tag{9}$$

(The neat look of the above explains the $\frac{1}{2}$ and $2B$ in (8).) Since $(0,0)$ is a critical point, (9) shows that $a = 0$ and $b = 0$, so our quadratic function has the form

$$w - w_0 = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2). \tag{10}$$

We moved w_0 to the left side since the tangent plane at $(0,0)$ is the horizontal plane $w = w_0$, and we are interested in whether the graph of the quadratic function lies above or below this tangent plane, i.e., whether $w - w_0 > 0$ or $w - w_0 < 0$ at points other than the origin.

If $(x,y) \neq (0,0)$, then either $x \neq 0$ or $y \neq 0$; say $y \neq 0$. Then we write (10) as

$$w - w_0 \;=\; \frac{y^2}{2}\left[A\left(\frac{x}{y}\right)^2 + 2B\left(\frac{x}{y}\right) + C\right] \tag{11}$$

We know that $y^2 > 0$ if $y \neq 0$; applying our previous lemma to the factor on the right of (11), (or if $y = 0$, switching the roles of x and y in (11) and applying the lemma), we get

$$\begin{array}{lll}
AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 & \Rightarrow & w - w_0 > 0 \quad \text{for all } (x,y) \neq (0,0); \\
& \Rightarrow & (0,0) \text{ is a minimum point;} \\
AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 & \Rightarrow & w - w_0 < 0 \quad \text{for all } (x,y) \neq (0,0); \\
& \Rightarrow & (0,0) \text{ is a maximum point;} \\
AC - B^2 < 0 & \Rightarrow & \begin{cases} w - w_0 > 0, & \text{for some } (x,y); \\ w - w_0 < 0, & \text{for some } (x,y); \end{cases} \\
& \Rightarrow & (0,0) \text{ is a saddle point.}
\end{array}$$

Argument for the Second-derivative Test for a general function.

This part won't be rigorous, only suggestive, but it will give the right idea.

We consider a general function $w = f(x, y)$, and assume it has a critical point at (x_0, y_0), and continuous second derivatives in the neighborhood of the critical point. Then by a generalization of Taylor's formula to functions of several variables, the function has a best quadratic approximation at the critical point. To simplify the notation, we will move the critical point to the origin by making the change of variables

$$u = x - x_0, \qquad v = y - y_0.$$

Then the best quadratic approximation is (if the x, y on the left and u, v on the right is upsetting, just imagine u and v replaced everywhere by $x - x_0$ and $y - y_0$):

$$w \;=\; f(x,y) \;\approx\; w_0 + \frac{1}{2}\big(Au^2 + 2Buv + Cv^2\big); \tag{13}$$

here the coefficients A, B, C are given as in (1) by the second partial derivatives with respect to u and v at $(0, 0)$, or what is the same (according to the chain rule—see the footnote below), by the second partial derivatives with respect to x and y at (x_0, y_0).

> (Intuitively, one can see the coefficients have these values by differentiating both sides of (13) and pretending the approximation is an equality. There are no linear terms in u and v on the right since $(0, 0)$ is a critical point.)

Since the quadratic function on the right of (13) is the best approximation to $w = f(x, y)$ for (x, y) close to (x_0, y_0), it is reasonable to suppose that their graphs are essentially the same near (x_0, y_0), so that if the quadratic function has a maximum, minimum or saddle point there, so will $f(x, y)$. Thus our results for the special case of a quadratic function having the origin as critical point carry over to the general function $f(x, y)$ at a critical point (x_0, y_0), if we interpret $A,\ B,\ C$ as the second partial derivatives at (x_0, y_0).

This is what the second derivative test says. □

Footnote: Using $u = x - x_0$ and $v = y - y_0$, we can apply the chain rule for partial derivatives, which tells us that for all x, y and the corresponding u, v, we have

$$w_x = w_u \frac{\partial u}{\partial x} + w_v \frac{\partial v}{\partial x} = w_u, \;\text{ since } u_x = 1 \text{ and } v_x = 0,$$

and similarly, $w_y = w_v$. Therefore at the corresponding points,

$$(w_x)_{(x_0,y_0)} = (w_u)_{(0,0)}, \qquad\qquad (w_y)_{(x_0,y_0)} = (w_v)_{(0,0)},$$

and differentiating once more and using the same reasoning,

$$(w_{xx})_{(x_0,y_0)} = (w_{uu})_{(0,0)}, \qquad (w_{xy})_{(x_0,y_0)} = (w_{uv})_{(0,0)}, \qquad (w_{yy})_{(x_0,y_0)} = (w_{vv})_{(0,0)}.$$

Second derivative test

1. Find and classify all the critical points of

$$w = (x^3 + 1)(y^3 + 1).$$

<u>**Answer:**</u> Taking the first partials and setting them to 0:

$$w_x = 3x^2(y^3 + 1) = 0 \quad \text{and} \quad w_y = 3y^2(x^3 + 1) = 0.$$

The first equation implies $x = 0$ or $y = -1$. We use these one at a time in the second equation.
If $x = 0$ then $w_y = 0 \Rightarrow y = 0 \Rightarrow (0, 0)$ is a critical point.
If $y = -1$ then $w_y = 0 \Rightarrow x^3 + 1 = 0 \Rightarrow x = -1 \Rightarrow (-1, -1)$ is a critical point.
The critical points are $(0, 0)$ and $(-1, -1)$.

Taking second partials:

$$w_{xx} = 6x(y^3 + 1), \quad w_{xy} = 9x^2y^2, \quad w_{yy} = 6y(x^3 + 1).$$

We analyze each critical point in turn.
At (-1,-1): $A = w_{xx}(-1, -1) = 0$, $B = w_{xy}(-1, -1) = 9$, $C = w_{yy}(-1, -1) = 0$. Therefore $AC - B^2 = -81 < 0$, which implies the critical point is a saddle.

At (0,0): $A = w_{xx}(0, 0) = 0$, $B = w_{xy}(0, 0) = 0$, $C = z_{yy}(-1, 2) = 0$. Therefore $AC - B^2 = 0$. The second derivative test fails.

Second derivative test

1. Find and classify all the critical points of

$$w = (x^3 + 1)(y^3 + 1).$$

Second derivative test

1. Find and classify all the critical points of

$$f(x, y) = x^6 + y^3 + 6x - 12y + 7.$$

<u>**Answer:**</u> Taking the first partials and setting them to 0:

$$\frac{\partial z}{\partial x} = 6x^5 + 6 = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 3y^2 - 12 = 0.$$

The first equation implies $x = -1$ and the second implies $y = \pm 2$. Thus, the critical points are $(-1, 2)$ and $(-1, -2)$.

Taking second partials:

$$\frac{\partial^2 z}{\partial x^2} = 30x^4, \quad \frac{\partial^2 z}{\partial xy} = 0, \quad \frac{\partial^2 z}{\partial y^2} = 6y.$$

We analyze each critical point in turn.
At (-1,-2): $A = z_{xx}(-1, -2) = 30$, $B = z_{xy}(-1, -2) = 0$, $C = z_{yy}(-1, -2) = -12$.
Therefore $AC - B^2 = -360 < 0$, which implies the critical point is a saddle.

At (-1,2): $A = z_{xx}(-1, 2) = 30$, $B = z_{xy}(-1, 2) = 0$, $C = z_{yy}(-1, 2) = 12$.
Therefore $AC - B^2 = 360 > 0$ and $A > 0$, which implies the critical point is a minimum.