Unit 3: Partial Derivatives

Dr.P.M.Bajracharya

June 3, 2022

Contents

1 Functions of several variables **8**

1.1 Functions of Two Variables 8

1.2 Graphing Functions of Two Variables 10

1.3 Level Curves 12

1.4 Functions of Three Variables 15

1.5 Functions of n Variables 17

2 Limits and continuity **18**

3 Partial derivatives **27**

3.1 Interpretations of Partial Derivatives 29

3.2 Functions of More Than Two Variables 30

3.3 Higher Derivatives 30

4 Partial Differential Equations **31**

4.1 Laplace's equation 31

4.2 Wave equation 32

4.3 The Cobb-Douglas Production Function 32

5 Tangent planes and linear approximation **36**

6 Chain rule **49**

7 Directional derivatives and gradient vector **54**

8 Maximum and minimum values **65**

9 Lagrange multipliers **83**

Multivariable Calculus in Data Science

You need to know some basic calculus in order to understand how functions change over time (derivatives) and to calculate the total amount of a quantity that accumulates over a time period (integrals). Other than that, Data Scientists mainly use calculus in building much Deep Learning and Machine Learning Models. They are involved in optimizing the data and bringing out better outputs of data, by drawing intelligent insights hidden in them.

Gradient Descent

A **gradient** measures how much the output of a function changes if you change the inputs a little bit.

Suppose you have a ball and a bowl. No matter wherever you slide the ball in the bowl, it will eventually land in the bottom of the bowl.

As you see this ball follows a path that ends at the bottom of the bowl. We can also say that the ball is descending in the bottom of the bowl. As you can see from the image the red lines are gradient of the bowl and the blue line is the path of the ball and as the path of the balls slope is decreasing, it is called as **gradient descent**.

In our machine learning model our goal is to reduce the cost in our input data. the cost function is used to monitor the error in predictions of an ML model. So minimizing this, basically means getting to the lowest error value possible or increasing the accuracy of the model. In short, We increase the accuracy by iterating over a training data set while tweaking the parameters(the weights and biases) of our model.

Lets us consider a dataset of users with their marks in some of the subjects and their occupation. Our goal is to predict the occupation of the person with considering the marks of the person.

	Math	Phy	Chem	Bio	Eng	Profession
John	78	56	65	83	66	Doctor
Eve	83	78	72	66	59	Engineer
Adam	75	67	79	55	70	?

In this dataset we have data of John and eve. With the reference data of john and eve, we have to predict the profession of Adam.

Now think of marks in the subject as a gradient and profession as the bottom target. You have to optimise your model so that the result it predicts at the bottom should be accurate. Using Johns and Eves data we will create gradient descent and tune our model such that if we enter the marks of john then it should predict result of Doctor in the bottom of gradient and same for Eve. This is our trained model. Now if we give marks of subject to our model then we can easily predict the profession.

gif2

In theory this is it for gradient descent, but to calculate and model, gradient descent requires calculus and now we can see importance of calculus in machine learning.

First Lets start by the topic that you know till now ie. Linear Algebra. Let first use linear algebra and its formula for our model.

gif3

The basic formula that we can use in this model is

$$y = mx + b,$$

where

$$y = \text{predictor}, \ m = \text{slope}, \ x = \text{input}, \ b = y\text{-intercept}.$$

A standard approach to solving this type of problem is to define an error function (also called a cost function) that measures how “good a given line is. This function will take in a (m, b) pair and return an error value based on how well the line fits our data. To compute this error for a given line, well iterate through each (x, y) point in our data set and sum the square distances between each points y-value and the candidate lines y-value (computed at $mx + b$). Its conventional to square this distance to ensure that it is positive and to make our error function differentiable.

$$\text{Error}_{(m,b)} = \frac{1}{N} \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

Lines that fit our data better (where better is defined by our error function) will result in lower error values. If we minimize this function, we will get the best line for our data. Since our error function consists of two parameters (m and b) we can visualize it as a two-dimensional surface. This is what it looks like for our data set:

gif4

Each point in this two-dimensional space represents a line. The height of the function at each point is the error value for that line. You can see that some lines yield smaller error values than others (i.e., fit our data better). When we run gradient descent search, we will start from some location on this surface and move downhill to find the line with the lowest error.

In the essence of calculus video you have seen that to calculate slope, we use differentiation.

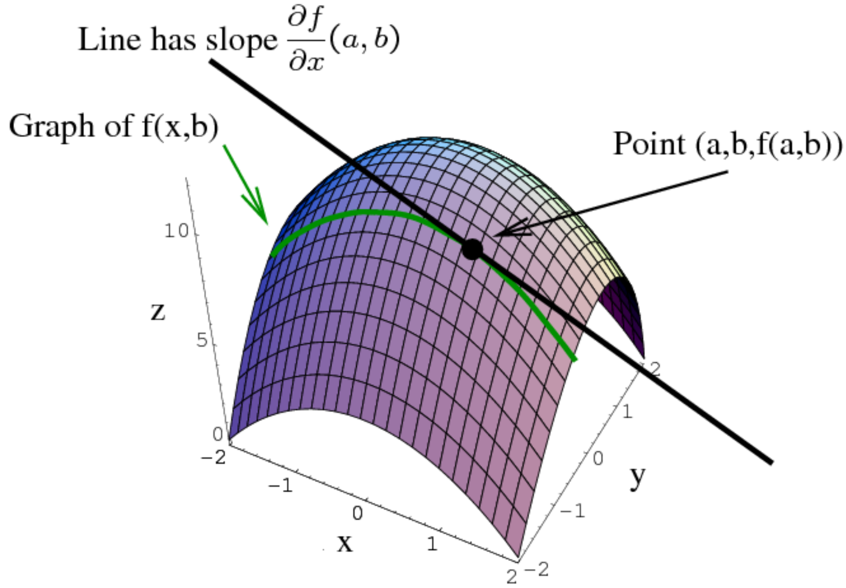


Figure 1: The graph of a function $z = f(x,y)z = f(x,y)$ is a surface, and fixing $y = by = b$ gives a curve (shown in green). The partial derivative $\partial f \partial x(a,b) \partial f \partial x(a,b)$ is the slope of the tangent line to this curve at the point where $x = ax = a$.

To run gradient descent on this error function, we first need to compute its gradient. The gradient will act like a compass and always point us downhill. To compute it, we will need to differentiate our error function. Since our function is defined by two parameters (m and b), we will need to compute a partial derivative for each. These derivatives work out to be:

$$\frac{\partial}{\partial m} = \frac{2}{N} \sum_{i=1}^{N} -x_i(y_i - (mx_i + b))$$

$$\frac{\partial}{\partial b} = \frac{2}{N} \sum_{i=1}^{N} -(y_i - (mx_i + b))$$

We now have all the tools needed to run gradient descent. We can initialize our search to start at any pair of m and b values (i.e., any line) and let the gradient descent algorithm march downhill on our error function towards the best line. Each iteration will update m and b to a line that yields slightly lower error than the previous iteration. The direction to move in for each iteration is calculated using the two partial derivatives from above.

The **Learning Rate** variable controls how large of a step we take downhill during each iteration. If we take too large of a step, we may step over the minimum. However, if we take small steps, it will require many iterations to arrive at the minimum.

While we were able to scratch the surface for learning gradient descent, there are several additional concepts that are good to be aware of that. A few of these include:

Convexity In our linear regression problem, there was only one minimum. Our error surface was convex. Regardless of where we started, we would eventually arrive at the absolute minimum. In general, this need not be the case. Its possible to have a problem with local minima that a gradient search can get stuck in. There are several approaches to mitigate this (e.g., stochastic gradient search).

gif5

Convergence We didnt talk about how to determine when the search finds a solution. This is typically done by looking for small changes in error iteration-to-iteration (e.g., where the gradient is near zero).

gif6

1 Functions of several variables

- Our first step is to explain what a function of more than one variable is, starting with functions of two independent variables. This step includes identifying
  - the domain and range of such functions and
  - learning how to graph them.
- We also examine ways to relate the graphs of functions in three dimensions to graphs of more familiar planar functions.

1.1 Functions of Two Variables

The definition of a function of two variables is very similar to the definition for a function of one variable. The main difference is that, instead of mapping values of one variable to values of another variable, we map ordered pairs of variables to another variable.

Definition.

Let $D \subseteq \mathbb{R}^2$. If to each ordered pair (x, y) in D there corresponds a unique real number $f(x, y)$ then f is called a function of x and y defined on D. The set D is called the **domain** of the function.
The **range** of f is the set of all real numbers z that has at least one ordered pair $(x, y) \in D$ such that $f(x, y) = z$ as shown in the following figure.

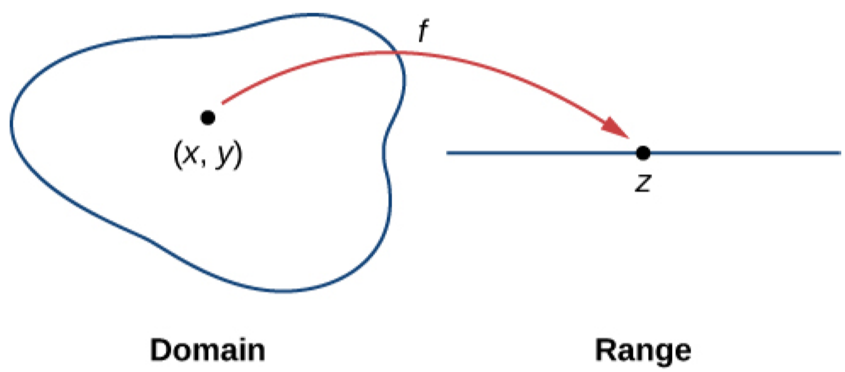


Figure 2: The domain of a function of two variables consists of ordered pairs (x, y).

Determining the domain of a function of two variables involves taking into account any domain restrictions that may exist. Let's take a look.

Example 1 (Domains and Range)**.** Find the domain and range of each of the following functions

(a) $f(x, y) = 3x + 5y + 2$

(b) $g(x, y) = \sqrt{9 - x^2 - y^2}$

Solution. a Domain $= \mathbb{R}^2$.

Range $= \mathbb{R}$.

b Domain $= \{(x, y) \in \mathbb{R}^2 |\ x^2 + y^2 \leq 9\}$ a disk of radius 3 centered at the origin including the boundary circle as shown in the following graph.

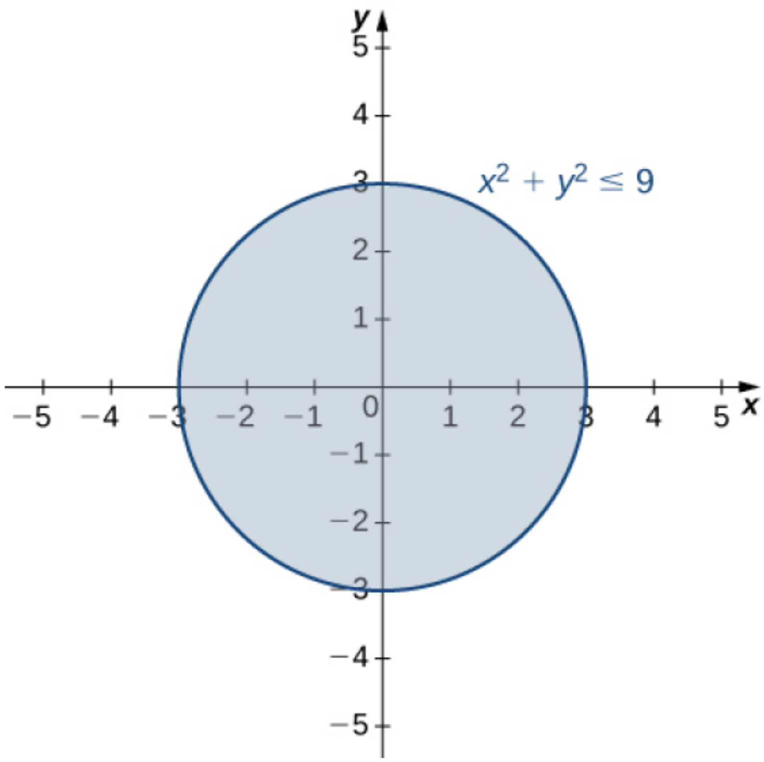


Figure 3: The domain of the function $g(x, y) = \sqrt{9 - x^2 - y^2}$ is a closed disk of radius 3.

Range $= [0, 3]$.

Problem 1. *Find the domain and range of the function* $f(x, y) = \sqrt{36 - 9x^2 - 9y^2}$

1.2 Graphing Functions of Two Variables

Suppose we wish to graph the function $z = f(x, y)$.

With a function of two variables, each ordered pair (x, y) in the domain of the function is mapped to a real number z. Therefore, the graph of the function f consists of ordered triples (x, y, z).

The graph of a function $z = f(x, y)$ of two variables is called a **surface**.

To understand more completely the concept of plotting a set of ordered triples to obtain a surface in three-dimensional space, imagine the (x, y) coordinate system laying flat. Then, every point in the domain of the function f has a unique z-value associated with it. If z is positive, then the graphed point is located above the xy-plane, if z is negative, then the graphed point is located below the xy-plane. The set of all the graphed points becomes the two-dimensional surface that is the graph of the function f .

Example 2. Sketch the graph of $g(x, y) = \sqrt{9 - x^2 - y^2}$.

Solution.

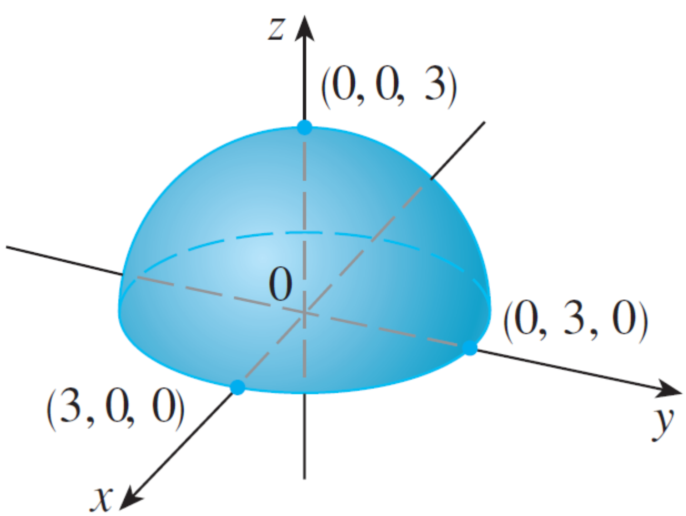


Figure 4: Graph of $g(x, y) = \sqrt{9 - x^2 - y^2}$

Problem 2. *Sketch the graph of $f(x, y) = x^2 + y^2$.*

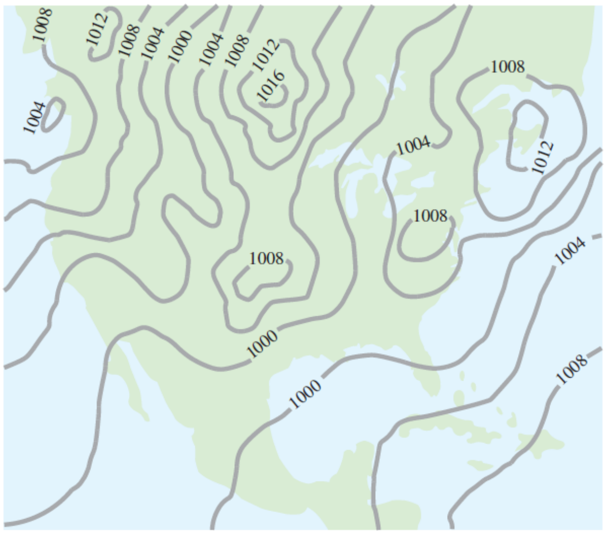
Problem 3. *A profit function for a hardware manufacturer is given by* $f(x, y) = 16 - (x - 3)^2 - (y - 2)^2$*, where* x *is the number of nuts sold per month (measured in thousands) and* y *represents the number of bolts sold per month (measured in thousands). Profit is measured in thousands of dollars. Sketch a graph of this function.*

1.3 Level Curves

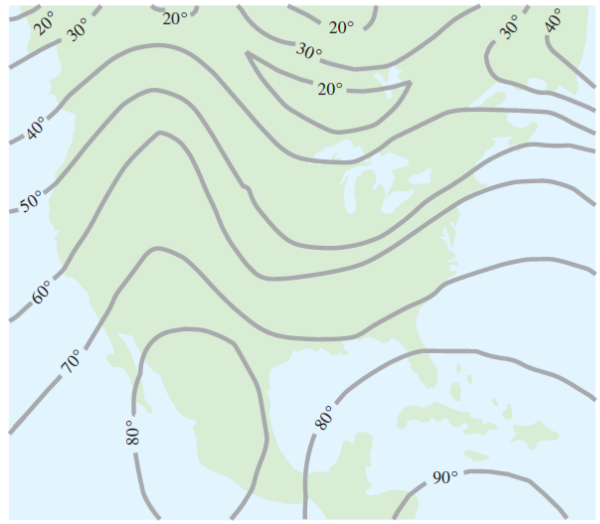
Another method for visualizing functions, borrowed from mapmakers, is a contour map on which points of constant elevation are joined to form *contour lines*, or *level curves*.

Definition.

The **level curves** of a function of two variables are the curves with equations $f(x, y) = c$, where c is a constant (in the range of f). A graph of the various level curves of a function is called a **contour map**.



Level curves show the lines of equal pressure (isobars), measured in millibars.



Level curves show the lines of equal temperature (isotherms), measured in degrees Fahrenheit.

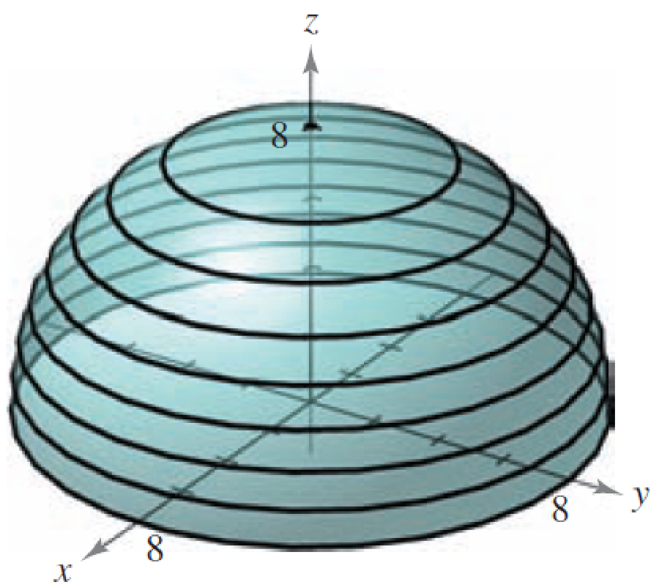


Figure 5: Level curves of a Hemisphere

Example 3. Create a contour map for the surface

$$g(x,y)=\sqrt{9-x^2-y^2}$$

corresponding to $c=0,1,2,3$.

Solution.

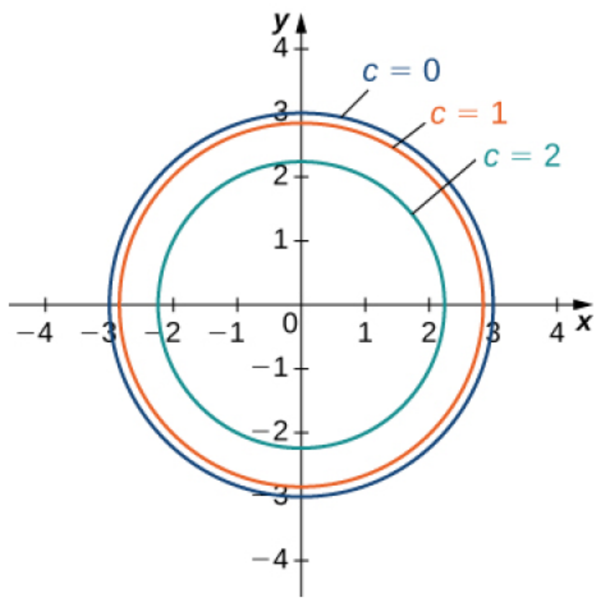


Figure 6: Level curves of the function
$g(x,y)=\sqrt{9-x^2-y^2}$, using $k=0,1,2,3$.

Note that in the previous derivation it may be possible that we introduced extra solutions by squaring both sides. This is not the case here because the range of the square root function is nonnegative.

Problem 4. *Sketch the graph of*

$$f(x, y) = \sqrt{64 - x^2 - y^2}.$$

Also, create a contour map for this surface corresponding to $c = 0, 1, 2, ..., 8$.

Problem 5. *Given the function*

$$f(x, y) = \sqrt{8 + 8x - 4y - 4x^2 - y^2},$$

find the level curve corresponding to $k = 0$. *Then create a contour map for this function. What are the domain and range of* f *?*

1.4 Functions of Three Variables

Let us take a brief look at functions of three variables. A function of three variables, f, is a rule that assigns to each ordered triple (x, y, z) in a domain $D \subseteq \mathbb{R}^3$ a unique real number denoted by $f(x, y, z)$.

For instance, the temperature T at a point on the surface of the earth depends on the longitude x and latitude y of the point and on the time t, so we could write $T = f(x, y, t)$.

Example 4. Find the domain of a function f given by

$$f(x, y, z) = \ln(z - y) + xy \sin z$$

Solution. $D = \{(x, y, z) \in \mathbb{R}^3 : \ z > y\}$.

This is a half-space consisting of all points that lie above the plane $z = y$.

Example 5. Find the domain of a function f given by

$$f(x, y, t) = \frac{\sqrt{2t - 4}}{x^2 - y^2}.$$

Solution. $D = \{(x, y, t) \in \mathbb{R}^3 : \ y \neq \pm x, t \geq 2\}$.

Problem 6. *Find the domain of each of the following functions:*

(a) $f(x, y, t) = (3t - 6)\sqrt{y - 4x^2 + 4}$

(b) $g(x, y, z) = \frac{3x - 4y + 2z}{\sqrt{9 - x^2 - y^2 - z^2}}$.

It's very difficult to visualize a function f of three variables by its graph, since that would lie in a four-dimensional space. However, we do gain some insight into f by examining its **level surfaces**, which are the surfaces with equations $f(x, y, z) = k$, where k is a constant. If the point (x, y, z) moves along a level surface, the value of $f(x, y, z)$ remains fixed.

Example 6. Find the level surfaces of the function

$$f(x, y, z) = x^2 + y^2 + z^2.$$

Solution. The level surfaces form a family of concentric spheres with radius $\sqrt{k}$.

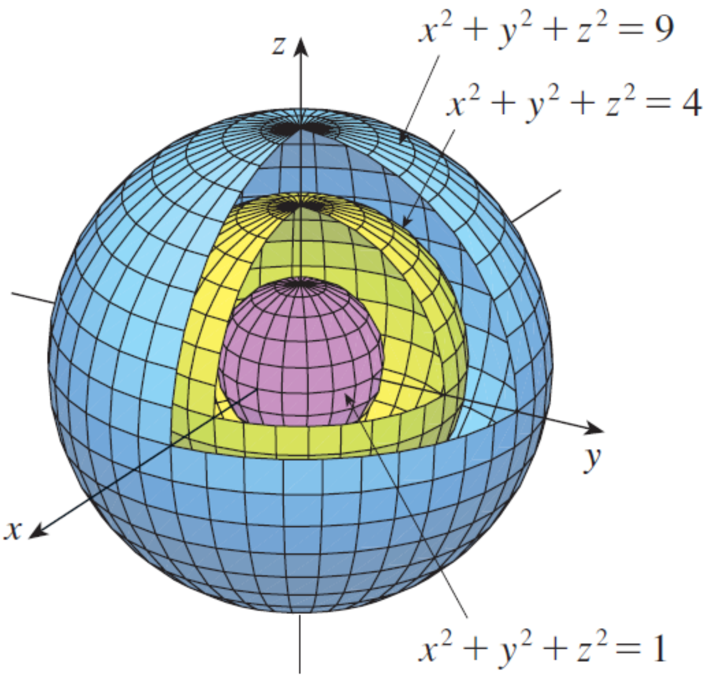


Figure 7: Level surfaces of a sphere.

Problem 7. *Find the level surface for the function and describe the surface, if possible.*

(a) $f(x, y, z) = 4x^2 + 9y^2 - z^2$ *corresponding to* $k = 1$*.*

(b) $f(x, y, z) = x^2 + y^2 + z^2 - 2x + 4y - 6z$ *corresponding to* $k = 2$*.*

1.5 Functions of n Variables

Functions of any number of variables can be considered. A function of n variables is a rule that assigns a number $z = f(x_1, x_2, ..., x_n)$ to a n-tuple of real numbers. We denote by $\mathbb{R}^n$ the set of all such n-tuples. For example, if a company uses n different ingredients in making a food product, c_i is the cost per unit of the ith ingredient, and units of the ingredient are used, then the total cost C of the ingredients is a function of the

n variables $x_1, x_2, ..., x_n$:

$$C = f(x_1, x_2, ..., x_n) = c_1 x_1 + c_2 x_2 + \cdot + c_n x_n.$$

The function f is a real-valued function whose domain is a subset of $\mathbb{R}^n$. Sometimes we will use vector notation to write such functions more compactly: If $x = (x_1, x_2, ..., x_n)$, we often write $f(x)$in place of $f(x_1, x_2, ..., x_n)$. With this notation we can rewrite the function defined in Equation 3 as

$$f(x) = c \cdot x,$$

where $c = (c_1, c_2, ..., c_n)$ and $c \cdot x$ denotes the dot product of the vectors c and x in V_n.

In view of the one-to-one correspondence between points $(x_1, x_2, ..., x_n)$ in $\mathbb{R}^n$ and their position vectors $x = (x_1, x_2, ..., x_n)$ in V_n, we have three ways of looking at a function f defined on a subset of $\mathbb{R}^n$:

1. As a function of n real variables $x_1, x_2, ..., x_n$.
2. As a function of a single point variable $(x_1, x_2, ..., x_n)$
3. As a function of a single vector variable $x = (x_1, x_2, ..., x_n)$

We will see that all three points of view are useful.

2 Limits and continuity

Limits

Let $u = (x, y) \in \mathbb{R}^2$. Then we write

$$\|u\| = \sqrt{x^2 + y^2}.$$

As you know, this is the Euclidean norm of u.

Let $D \subseteq \mathbb{R}^2$ and $p = (a, b) \in \mathbb{R}^2$. The point p is called a limitpoint or accumulation point of D if D includes points arbitrarily close to p, i.e.,

$$\forall\, r > 0\ \exists\, x \in D :\ 0 < \|p - x\| < r.$$

Definition. Let f be a real valued function defined on $D \subseteq \mathbb{R}^2$ and $p = (a, b) \in \mathbb{R}^2$ be a limitpoint of D. Then we say that L is the limit of $f(u)$ as $u = (x, y)$ tends to p if for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that

$$x, y \in D,\ 0 < \|x - p\| < \delta \Rightarrow |f(x, y) - L| < \varepsilon.$$

In this case, we write

$$\lim_{u \to p} f(u) = L.$$

Other notations for the limit in the above definition are

$$\lim_{(x,y)\to(a,b)} f(x, y) = L \quad \text{and} \qquad f(x, y) \to L \text{ as } (x, y) \to (a, b).$$

The above definition says that the distance between $f(u)$ and L can be made arbitrarily small by making the distance from u to p sufficiently small (but not 0). The definition refers only to the distance between u and p. It does not refer to the direction of approach. Therefore, if the limit exists, then $f(u)$ must approach the same limit no matter how u approaches p. Thus, if we can find two different paths of approach along which the function $f(u)$ has different limits, then it follows that $\lim_{u\to p} f(u)$ does not exist.

Let $f(u) \to L_1$ as $u \to p$ along a path C_1 and $f(u) \to L_2$ as $u \to p$ along a path C_2. If $L_1 \neq L_2$, then the limit $\lim_{u \to p} f(u)$ does not exist.

Example 7. Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ does not exist.

Solution. Let's approach $(0, 0)$ first along the x-axis and then along the y-axis.

Example 8. If $f(x, y) = \dfrac{xy}{x^2 + y^2}$, does the limit $\displaystyle\lim_{(x,y)\to(0,0)} f(x, y)$ exist?

Solution. Let's approach $(0, 0)$ first along the x-axis and along the y-axis. Then approach $(0, 0)$ along another line, say, $x = y$ for all $x \neq 0$. No!

Example 9. If $f(x, y) = \dfrac{xy^2}{x^2 + y^4}$, does the limit $\displaystyle\lim_{(x,y)\to(0,0)} f(x, y)$ exist?

Solution. Let's approach $(0, 0)$ along any nonvertical line $y = mx$ through the origin. But if we approach $(0, 0)$ Then approach $(0, 0)$ along along the parabola $x = y^2$, then $f(x, y) = 1/2$. No!

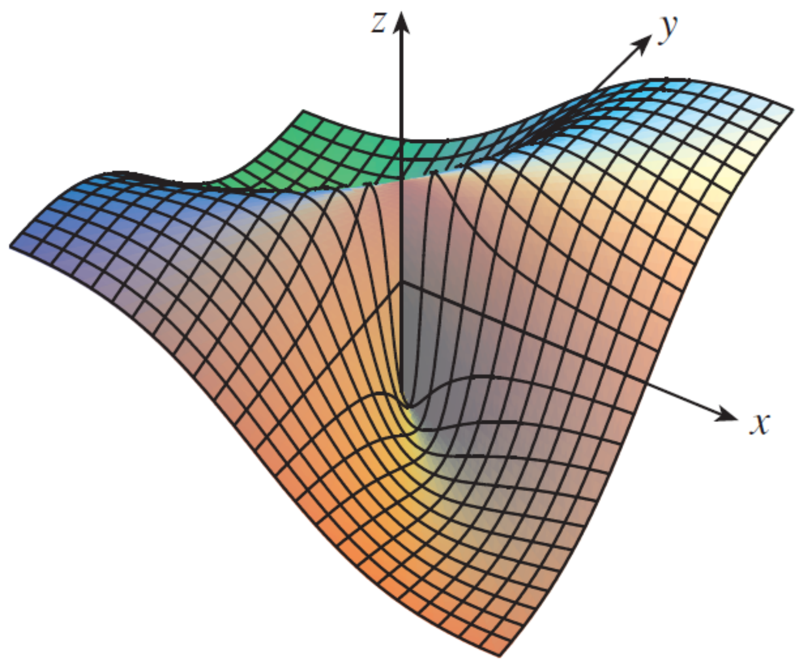


Figure 8

Limit laws:

Let $u = (x, y)$ and $p = (a, b)$. If $\lim_{u\to p} f(u) = L$, $\lim_{u\to p} g(u) = M$, then the following sum, product, and quotient rules, and squeeze theorem hold.

(a) $\lim_{u\to p}(f(u) + g(u)) = L + M$

(b) $\lim_{u\to p}(f(u)g(u)) = LM$

(c) $\lim_{u\to p} \dfrac{f(u)}{g(u)} = \dfrac{L}{M}$ $(M \neq 0)$.

Squeeze theorem: Let

$$\lim_{u\to p} f(u) = L, \quad \lim_{u\to p} g(u) = M.$$

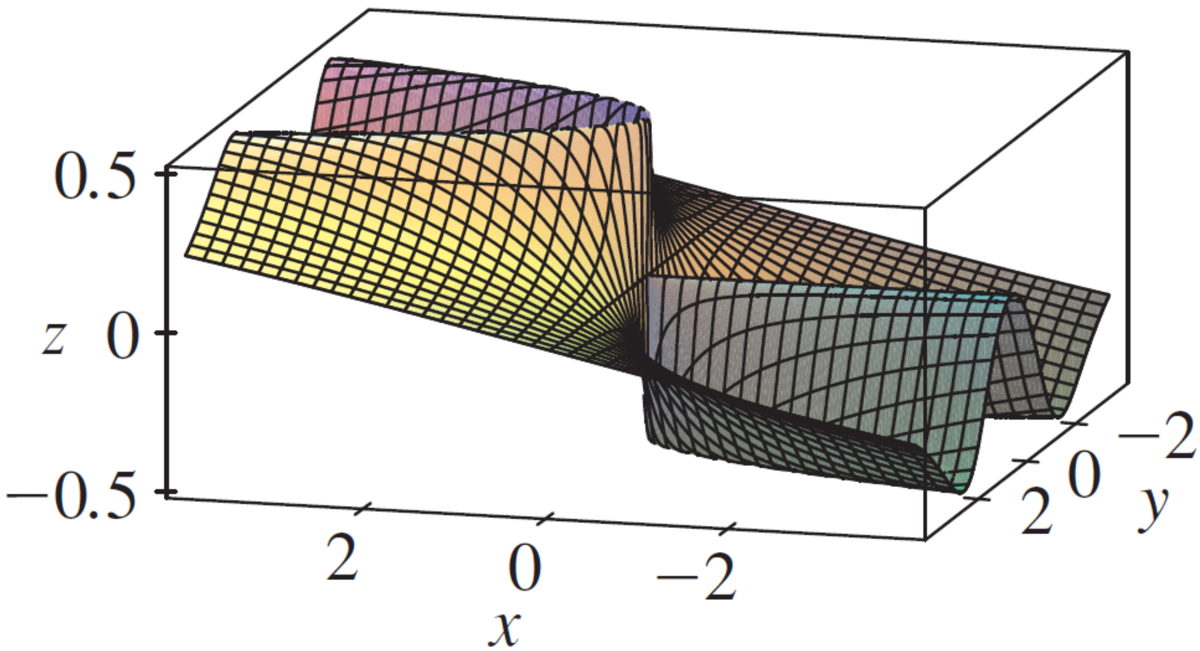


Figure 9

If

$$\lim_{u\to p} f(u) = \lim_{u\to p} g(u) \text{ and } f(u) \le h(u) \le g(u),$$

then $\lim_{u\to p} h(u)$ exists and equals L which equals M.

Guessing and proving a limit

Example 10. Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2}$, if exists.

Solution. As in the last example, we could show that the limit along any line through the origin is 0. The limits along the parabolas $y = x^2$ and $x = y^2$ also turn out to be 0, so we begin to suspect that the limit does exist and is equal to 0. Let's prove it using the Squeeze Theorem.

Continuity

A function f of two variables is called **continuous at a point** (a, b) in a set $D \subseteq \mathbb{R}^2$ if the following conditions are satisfied:

1. $f(a, b)$ exists.
2. $\lim\limits_{(x,y)\to(a,b)} f(x, y)$ exists.
3. $\lim\limits_{(x,y)\to(a,b)} f(x, y) = f(a, b)$.

We say f is **continuous on a set** $D \subseteq \mathbb{R}^2$ if f is continuous at every point (a, b) in D.

The intuitive meaning of continuity is that if the point (x, y) changes by a small amount, then the value of $f(x, y)$ changes by a small amount. This means that a surface that is the graph of a continuous function has no hole or break.

Using the properties of limits, you can see that sums, differences, products, and quotients of continuous functions are continuous on their domains. Moreover, the following property related to a composition of two functions also holds:

Let $\lim\limits_{(x,y)\to(a,b)} f(x,y) = L$. If $F(t)$ is a continuous function at $t = L$, then

$$\lim_{(x,y)\to(a,b)} F(f(x,y)) = F(L) = F(\lim_{(x,y)\to(a,b)} f(x,y)).$$

That is, for continuous functions, we may interchange the limit and function composition operations.

Let's use these facts to give examples of continuous functions.

It is easy to show that

$$\lim_{(x,y)\to(a,b)} x = a, \qquad \lim_{(x,y)\to(a,b)} y = b, \qquad \lim_{(x,y)\to(a,b)} c = c.$$

These limits show that the functions $f(x,y) = x$, $g(x,y) = y$, and $h(x,y) = c$ are continuous everywhere on $\mathbb{R}^2$.

A **polynomial function of two variables** is a sum of terms of the form $cx^m y^n$, where c is a constant and m and n are nonnegative integers. It follows that all polynomials are continuous on $\mathbb{R}^2$.

Likewise, any rational function $f(x,y) = \dfrac{P(x,y)}{Q(x,y)}$, where $P(x,y), Q(x,y)$ are polynomials and $Q(x,y) \neq 0$ is continuous on its domain because it is a quotient of continuous functions $P(x,y)$ and $Q(x,y)$.

Example 11. Evaluate $\lim\limits_{(x,y)\to(a,b)} (x^2y^3 - x^3y^2 + 3x + 2y)$.

Solution.

Example 12. Where is the function $f(x,y) = \dfrac{x^2 - y^2}{x^2 + y^2}$ continuous?

Solution.

Example 13. Show that the function

$$f(x,y) = \begin{cases} \dfrac{x^2 - y^2}{x^2 + y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

is discontinuous at the origin.

Solution.

Example 14. Show that the function

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

is continuous everywhere on $\mathbb{R}^2$.

Solution.

Example 15. Where is the function $h(x,y) = \arctan(y/x)$ continuous?

Solution. The function $f(x,y) = y/x$ is a rational function and therefore continuous except on the line $x = 0$. The function $g(t) = \arctan t$ is continuous everywhere. So the composite function

$$g(f(x,y)) = \arctan(y/x) = h(x,y)$$

is continuous except where $x = 0$. The graph in the following figure shows the break in the graph of h above the x-axis.

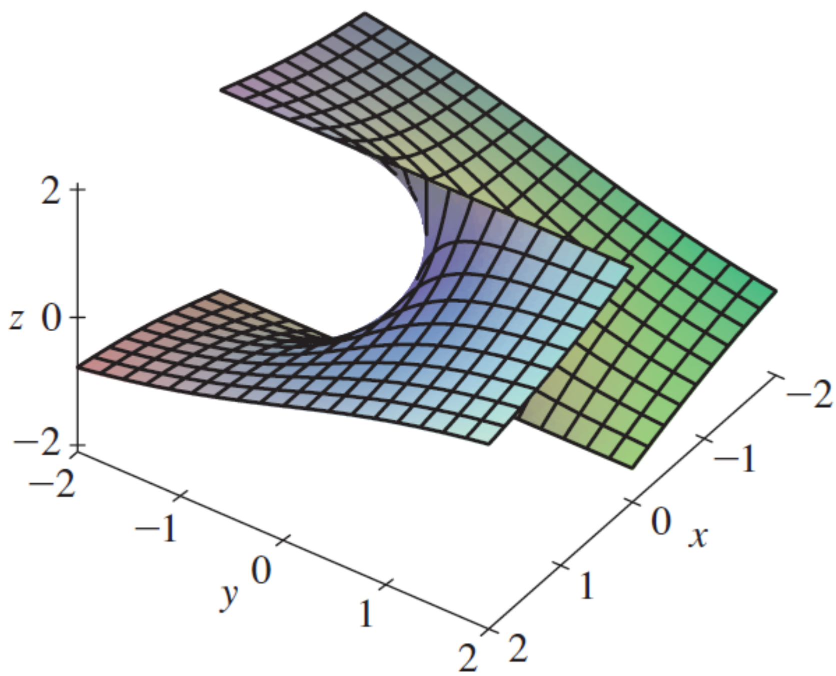


Figure 10: The function $h(x,y) = arctan(y/x)$
is discontinuous where $x = 0$.

Everything that we have done in this section can be extended to functions of three or more variables. The notation

$$\lim_{(x,y,z)\to(a,b,c)} f(x,y,z) = L$$

means that the values of $f(x,y,z)$ approach the number as the point (x,y,z) approaches the point (a,b,c) along any path in the domain of f. The function f is continuous at (a, b, c) if

$$\lim_{(x,y,z)\to(a,b,c)} f(x, y, z) = f(a, b, c)$$

For instance, the function

$$f(x, y, z) = \frac{1}{x^2 + y^2 + z^2 - 1}$$

is a rational function of three variables and so is continuous at every point in $\mathbb{R}^2$ except where $x^2+y^2+z^2 = 1$. In other words, it is discontinuous on the sphere with center at the origin and of radius 1.

3 Partial derivatives

It is worthwhile to note that a function f of two or more variables does not have a unique rate of change because each variable may affect f in different ways.

For example, the current I in a circuit is a function of both voltage V and resistance R given by Ohm's Law:

$$I(V, R) = \frac{V}{R}.$$

The current I is increasing as a function of V but decreasing as a function of R.

The partial derivatives are the rates of change with respect to each variable separately. A function $f(x, y)$ of two variables has two partial derivatives, denoted f_x and f_y , defined by the following limits (if they exist):

$$f_x(a,b) = \lim_{h\to 0} \frac{f(a+h,b)-f(a,b)}{h}$$
$$f_y(a,b) = \lim_{k\to 0} \frac{f(a,b+k)-f(a,b)}{k}$$

Thus, f_x is the derivative of $f(x,b)$ as a function of x alone, and f_y is the derivative of $f(a,y)$ as a function of y alone. The Leibniz notation for partial derivatives is

$$\frac{\partial f}{\partial x} = D_x f = f_x, \qquad \frac{\partial f}{\partial y} = D_y f = f_y,$$

$$\left.\frac{\partial f}{\partial x}\right|_{(a,b)} = f_x(a,b), \qquad \left.\frac{\partial f}{\partial y}\right|_{(a,b)} = f_y(a,b).$$

To compute partial derivatives, all we have to do is remember that the partial derivative with respect to x is just the ordinary derivative of the function g of a single variable that we get by keeping y fixed. Thus we have the following rule.

Rule for Finding Partial Derivatives of $z = f(x,y)$

1. To find f_x, regard y as a constant and differentiate $f(x,y)$ with respect to x.
2. To find f_y, regard x as a constant and differentiate $f(x,y)$ with respect to y.

3.1 Interpretations of Partial Derivatives

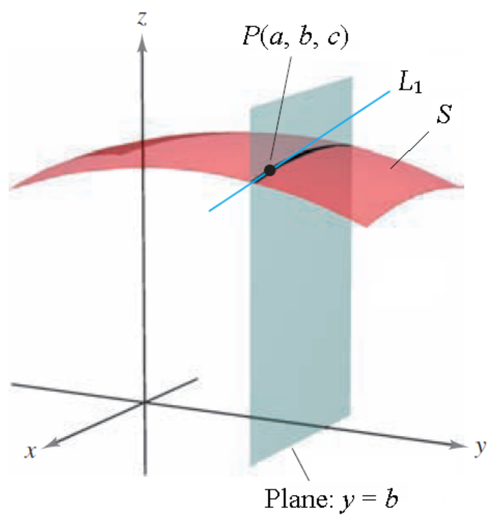


Figure 11: $f_x(a, b)$ = slope in x-direction.

$\left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)}$ – the slope of the tangent line L_1 to the curve $f(x, y_0)$ at (x_0, y_0).

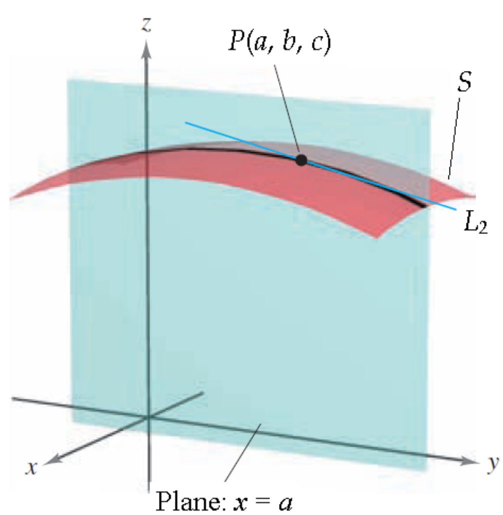


Figure 12: $f_y(a, b)$ = slope in y-direction.

$\left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)}$ – the slope of the tangent line L_2 to the curve $f(x_0, y)$ at (x_0, y_0).

Informally, the values of f_x and f_y at the point denote the slopes of the surface in the x- and y-directions, respectively.

Example 16. For $f(x,y) = xe^{x^2y}$, find f_x and f_y, and evaluate each at the point $(1, \ln 2)$.

Solution.

Example 17 (Implicit partial differentiation:)**.** Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ if z is defined implicitly as a function of x and y by the equation

$$x^3 + y^3 + z^3 + 6xyz = 1.$$

Solution.

3.2 Functions of More Than Two Variables

3.3 Higher Derivatives

Example 18. Find the second partial derivatives of

$$f(x,y) = x^3 + x^2y^3 + 2y^2.$$

Solution.

The following theorem, which was discovered by the French mathematician Alexis Clairaut (1713-1765), gives conditions under which we can assert that $f_{xy} = f_{yx}$.

Theorem 3.1. *Suppose f is defined on a disk D that contains the point (a, b). If the functions f_{xy} and f_{yx} are both continuous on D, then*

$$f_{xy}(a, b) = f_{yx}(a, b).$$

Proof.

◄

Partial derivatives of order 3 or higher

Example 19. Calculate f_{xxyz} if $f(x, y, z) = \sin(3x + yz)$.

4 Partial Differential Equations

A **partial differential equation** (PDE) is a differential equation involving functions of several variables and their partial derivatives.

4.1 Laplace's equation

The partial differential equation of the form:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

is called **Laplace's equation**. Solutions of this equation are called **harmonic functions**. They play a role in problems of heat conduction, fluid flow, and electric potential.

Example 20. Show that the function $u(x, y) = e^x \sin y$ is a solution of Laplace's equation.

Solution.

4.2 Wave equation

The partial differential equation of the form:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$$

is called a **wave equation**.

This equation describes the motion of a waveform, which could be an ocean wave, a sound wave, a light wave, or a wave traveling along a vibrating string. For instance, if $u(x, t)$ represents the displacement of a vibrating violin string at time t and at a distance x from one end of the string (as in Figure 8), then $u(x, t)$ satisfies the wave equation. Here the constant depends on the density of the string and on the tension in the string.

Example 21. Verify that the function $u(x, y) = \sin(x - at)$ satisfies the wave equation.

Solution.

4.3 The Cobb-Douglas Production Function

Let

P : total production of an economic system

L : the amount of labor required to produce P

K : the capital investment required to produce P.

Then the total production P can be described as a function of L and K.

If the production function is denoted by $P = P(L, K)$, then the partial derivative $\partial P/\partial L$ is the rate at which production changes with respect to the amount of labor. Economists call it the **marginal production with respect to labor** or the **marginal productivity of labor**.

Likewise, the partial derivative $\partial P/\partial K$ is the rate of change of production with respect to capital and is called the **marginal productivity of capital**. In these terms, the assumptions made by Cobb and Douglas can be stated as follows.

(i) If either labor or capital vanishes, then so will production.

(ii) The marginal productivity of labor is proportional to the amount of production per unit of labor.

(iii) The marginal productivity of capital is proportional to the amount of production per unit of capital.

Because the production per unit of labor is P/L, assumption (ii) says that

$$\frac{\partial P}{\partial L} = \alpha \frac{P}{L}$$

for some constant α. If we keep K constant , then this partial differential equation becomes an ordinary differential equation:

$$\frac{dP}{dL} = \alpha \frac{P}{L}$$

If we solve this separable differential equation, we get

$$P(L, K_0) = C_1(K_0)L^{\alpha} \tag{1}$$

Notice that we have written the constant C_1 as a function of K_0 because it could depend on the value of K_0.

Similarly, assumption (iii) says that

$$\frac{\partial P}{\partial K} = \beta \frac{P}{K}$$

and we can solve this differential equation to get

$$P(L_0, K) = C_1(L_0)K^{\beta} \tag{2}$$

Comparing Equations (1) and (2), we have

$$P(L, K) = bL^{\alpha}K^{\beta}, \tag{3}$$

where b is a constant that is independent of both L and K. Assumption (i) shows that $\alpha > 0$ and $\beta > 0$.

Notice from Equation (3) that if labor and capital are both increased by a factor m, then

$$P(mL, mK) = b(mL)^{\alpha}(mK)^{\beta} = m^{\alpha+\beta}P(L, K)$$

If $\alpha + \beta = 1$, then $P(mL, mK) = mP(L, K)$, which means that production is also increased by a factor of m. That is why Cobb and Douglas assumed that $\alpha + \beta = 1$ and therefore

$$P(L, K) = bL^{\alpha}K^{1-\alpha},$$

This is the Cobb-Douglas production function.

Example 22. Consider the Cobb-Douglas production model given by the formula $P = L^{0.75}K^{0.25}$. Sketch the level curves

$$P(L, K) = 1, P(L, K) = 2, P(L, K) = 3$$

in an LK-coordinate system (L horizontal and K vertical).

Example 23. Show that the Cobb-Douglas production function $P = bL^{\alpha}K^{\beta}$ satisfies the equation

$$L\frac{\partial P}{\partial L} + K\frac{\partial P}{\partial K} = (\alpha + \beta)P.$$

Note that if $\alpha+\beta = 1$, then the production function P satisfies the differential equation:

$$L\frac{\partial P}{\partial L} + K\frac{\partial P}{\partial K} = P.$$

Example 24. Let's consider a small printing business where N is the number of workers, V is the value of the equipment (in units of \$25,000), and P is the production, measured in thousands of pages per day. Suppose the production function for this company is given by

$$P = f(N, V) = 2N^{0.6}V^{0.4}.$$

(a) If this company has a labor force of 100 workers and 200 units' worth of equipment, what is the production output of the company?

(b) Find $f_N(100, 200)$ and $f_V(100, 200)$. Interpret your answers in terms of production.

5 Tangent planes and linear approximation

- Determine the equation of a plane tangent to a given surface at a point.
- Use the tangent plane to approximate a function of two variables at a point.
- Explain when a function of two variables is differentiable.
- Use the total differential to approximate the change in a function of two variables.

Tangent Planes

A function of one variable: $y = f(x)$.

The slope of the tangent line at the point $x = a$: $m = f'(a)$.

The equation of the tangent line at the point $x = a$:

$$y = f(a) + f'(a)(x - a).$$

What is the slope of a tangent plane?

Definition. Let $P_0 = (x_0, y_0, z_0)$ be a point on a surface S, and let C be any curve passing through P_0 and lying entirely in S. If the tangent lines to all such curves C at P_0 lie in the same plane, then this plane is called the **tangent plane** to S at P_0.

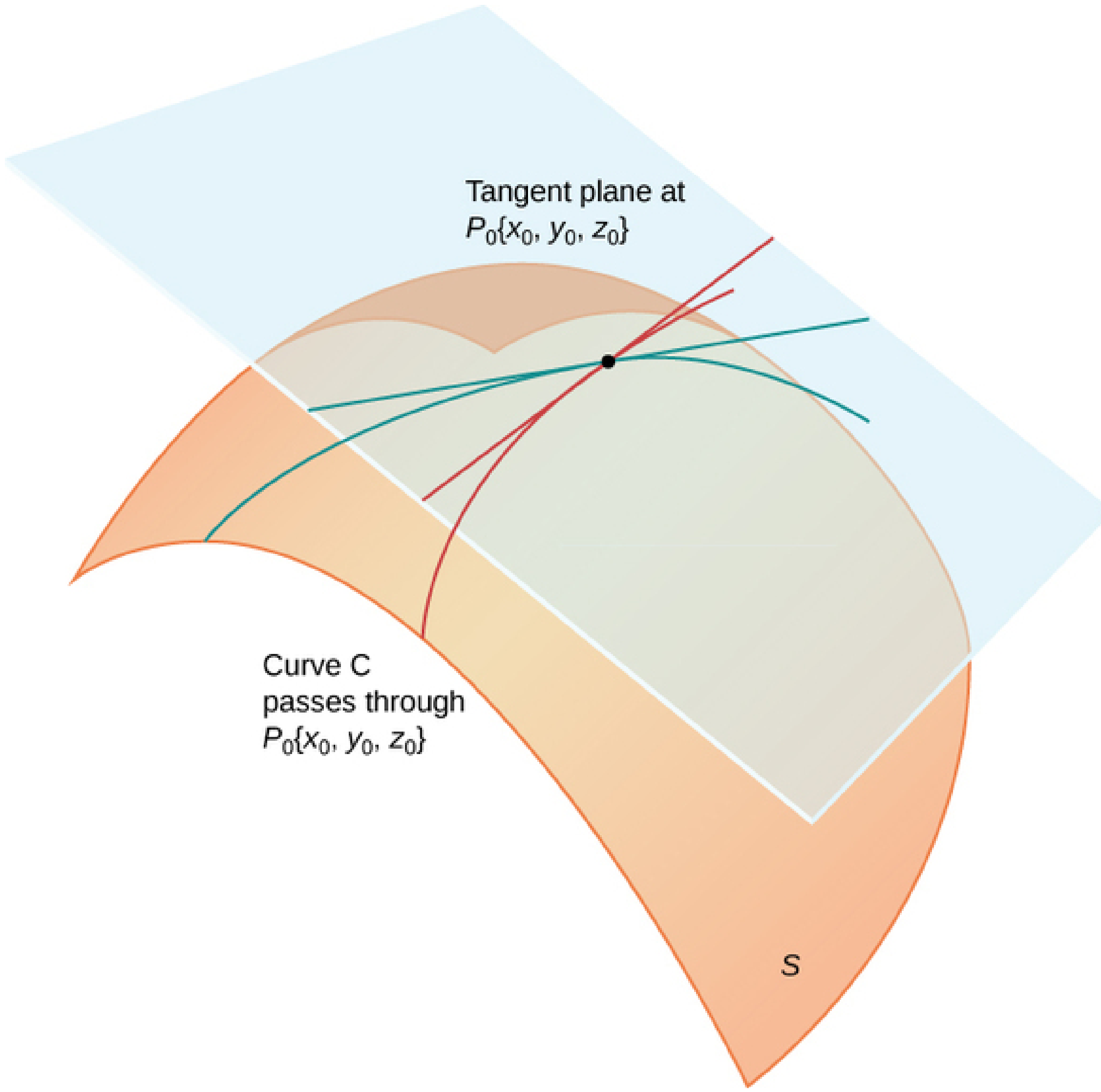


Figure 13: The tangent plane to a surface at a point contains all the tangent lines to curves in that pass through.

Equation of a tangent plane

We know that any plane passing through the point (x_0, y_0, z_0) has an equation of the form

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{4}$$

Dividing this equation by C, we can write it in the form

$$z - z_0 = a(x - x_0) + b(y - y_0). \tag{5}$$

Definition. Let S be a surface defined by a function $z = f(x, y)$ and $P_0 = (x_0, y_0)$ a point in the domain of f. Suppose that f has continuous partial derivatives. Then, the **equation of the tangent plane** to S at P_0 is given by

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \tag{6}$$

Example 25. Find the tangent plane to the elliptic paraboloid $z = 2x^2 + y^2$ at the point $(1, 1, 3)$.

Solution.

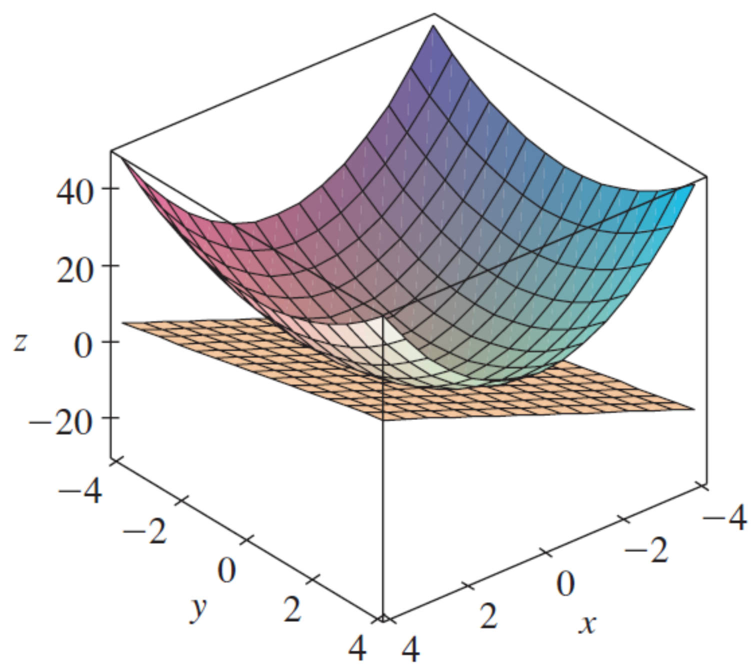


Figure 14: The tangent plane to $z = 2x^2 + y^2$ at the point $(1, 1, 3)$.

Linear Approximations

One nice use of tangent planes is they give us a way to approximate a surface near a point. As long as we are near to the point (x_0, y_0) then the tangent plane should nearly approximate the function at that point.

In Example 25, we found that an equation of the tangent plane to the graph of the function $f(x, y) = 2x^2 + y^2$ at the point $(1, 1, 3)$ is

$$z = 4x + 2y - 3.$$

Set

$$L(x, y) = 4x + 2y - 3.$$

This is a linear function in two variables

For instance, at the point $(1.1, 0.95)$ we have

$$f(1.1, 0.95) = 2(1.1)^2 + (0.95)^2 = 3.3225,$$
$$L(1.1, 0.95) = 4(1.1) + 2(0.95) - 3 = 3.3$$

Clearly, $f(x, y) \approx L(x, y)$. But if we take a point farther away from $(1, 1)$, such as $(2, 3)$, we no longer get a good approximation. In fact,

$$f(2, 3) = 2(2)^2 + (3)^2 = 17,$$
$$L(2, 3) = 4(2) + 2(3) - 3 = 11.$$

Thus, $L(x, y)$ is a good approximation to $f(x, y)$ when (x, y) is near $(1,\ 1)$. The function L is called the **linearization** of f at $(1, 1)$ and the approximation

$$f(x, y) \approx 4x + 2y - 3$$

is called the **linear approximation** or **tangent plane approximation** of f at $(1,1)$.

Because of this we define the linear approximation to be,

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

In general, we know that an equation of the tangent plane to the graph of a function $f(x,y)$ at the point $(a,b,f(a,b))$ is

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

The linear function

$$L(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

is called the **linearization**of f at (a,b) and the approximation

$$f(x,y) \approx L(x,y).$$

is called the **linear approximation** or the **tangent plane approximation** of f at (a,b).

We have defined tangent planes for surfaces $z = f(x,y)$, where f has continuous first partial derivatives. What happens if f_x and f_y are not continuous?

Example 26. Consider the following function:

$$f(x,y)=\begin{cases}\dfrac{xy}{x^2+y^2} & \text{if } (x,y)\neq(0,0)\\ 0 & \text{if } (x,y)=(0,0).\end{cases}$$

We can verify that its partial derivatives exist at the origin and, in fact, $f_x(0,0)=f_y(0,0)=0$, but f_x and f_y are not continuous. The linear approximation would be $f(x,y)\approx 0$, but $f(x,y)=\frac{1}{2}$ at all points on the line $y=x$. So a function of two variables can behave badly even though both of its partial derivatives exist.

To rule out such behavior as in the above example, we formulate the idea of a differentiable function of two variables.

Differentiable functions of two variables

Let us return to the one-dimensional case. If Δx is an increment in x, then the increment in y, Δy, is defined as

$$\Delta y = f(x+\Delta x)-f(x).$$

By definition, the derivative of $f:\mathbb{R}\to\mathbb{R}$ is given by

$$f'(x)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}.$$

This is the same as

$$\lim_{\Delta x\to 0}\frac{\Delta y - f'(x)\Delta x}{\Delta x}=0.$$

This gives

$$\Delta y = f'(x)\Delta x+\varepsilon\Delta x,\quad \text{where } \varepsilon\to 0 \text{ as } \Delta x\to 0. \tag{7}$$

Now consider a function of two variables, $z = f(x, y)$, and suppose that Δx is an increment in x and Δy an increment in y. Then the corresponding increment of z is given by

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$$

By analogy with (7) we define the differentiability of a function of two variables as follows.

Definition. Let $z = f(x, y)$. We say that f is differentiable at (a, b) if Δz can be expressed in the form

$$\Delta z = f_x(a, b)\Delta x + f_y(a, b)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $(\Delta x, \Delta y) \to (0, 0)$.

The above definition says that a differentiable function f is one for which the linear approximation

$$\begin{aligned} f(x, y) &\approx L(x, y) \\ &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \end{aligned}$$

is a good approximation when (x, y) is near (a, b). In other words, the tangent plane approximates the graph of f well near the point of tangency.

It's sometimes hard to use above definition directly to check the differentiability of a function, but the next theorem provides a convenient sufficient condition for differentiability.

Theorem 5.1. *If the partial derivatives f_x and f_y exist near (a, b) and are continuous at (a, b), then f is differentiable at (a, b).*

Example 27 (Using a linearization to estimate a function value)**.** Show that $f(x,y) = xe^{xy}$ is differentiable at $(1,0)$ and find its linearization there. Then use it to approximate $f(1.1, -0.1)$.

Solution. The partial derivatives are

$$f_x(x,y) = e^{xy} + xye^{xy}, \qquad f_y(x,y) = x^2e^{xy}.$$

This implies that

$$f_x(1,0) = 1, \qquad f_y(1,0) = 1.$$

Both f_x and f_y are continuous functions, so f is differentiable. The linearization is

$$\begin{aligned} L(x,y) &= f(1,0) + f_x(1,0)(x-1) + f_y(1,0)(y-0) \\ &= 1 + 1(1-0) + 1 \cdot y = x + y. \end{aligned}$$

The corresponding linear approximation is

$$xe^{xy} \approx x + y$$

and so

$$f(1.1, -0.1) \approx 1.1 - 0.1 = 1.$$

Compare this with the actual value of

$$f(1.1, -0.1) = 1.1e^{-0.11} \approx 0.98542.$$

Differentials

Recall: one dimensional case.

For a differentiable function of two variables, $z = f(x, y)$, we define the differentials dx and dy to be independent variables; that is, they can be given any values.

The differential dz, also called the **total differential**, is defined by

$$dz = f_x(x, y)dx + f_y(x, y)dy = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy.$$

If we take

$$dx = \Delta x = x - a, \quad dy = \Delta y = y - b,$$

then the differential dz is

$$dz = f_x(x, y)(x - a) + f_y(x, y)(y - b).$$

So, the linear approximation can be written as

$$f(x, y) \approx f(a, b) + dz.$$

Figure 7 is the three-dimensional counterpart of Figure 6 and shows the geometric interpretation of the differential and the increment : represents the change in height of the tangent plane, whereas represents the change in height of the surface when changes from to .

Example 28 (Differentials versus increments)**.**

(a) If $z = f(x, y) = x^2 + 3xy - y^2$, find the differential dz.

(b) If x changes from 2 to 2.05 and y changes from 3 to 2.96, compare the values of Δz and dz.

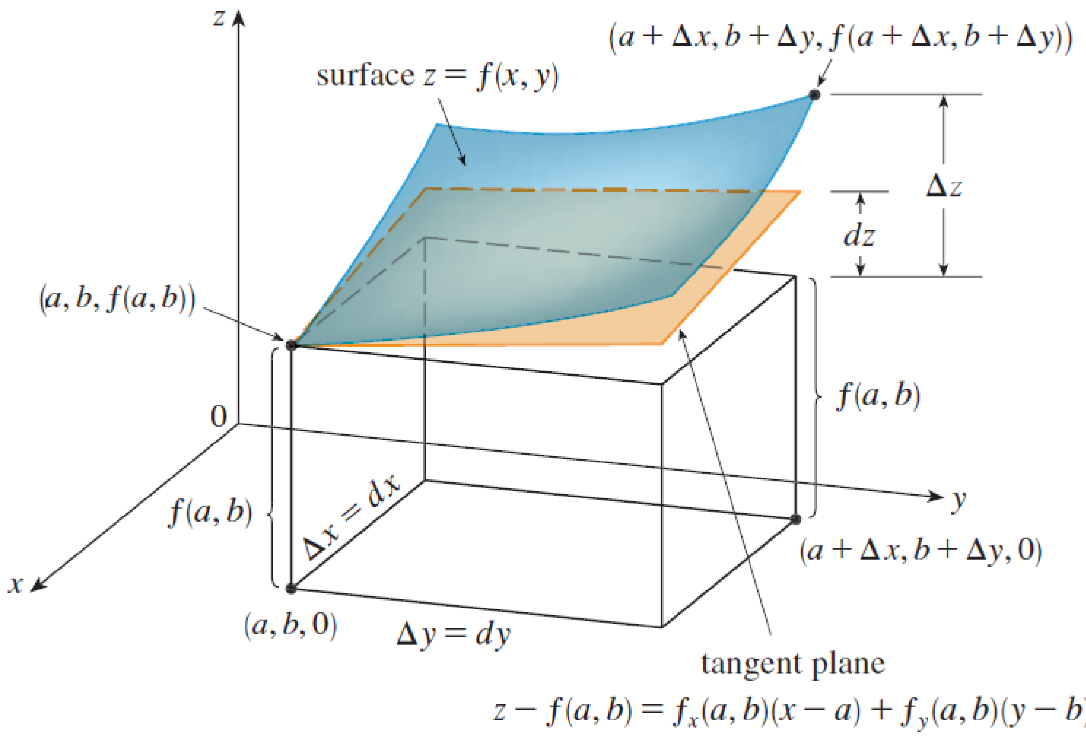


Figure 15

Solution. (a)

(b) Putting

$$x = 2, dx = \Delta x = 0.05, \quad y = 3, \quad dy = \Delta y = -0.04,$$

we get

$$dz = 2(2) + 3(3)0.05[3(2) - 2(3)](-0.04) = 0.65$$

The increment of z is

$$\begin{aligned}\Delta z &= f(2.05, 2.96) - f(2,3)\\ &= [2.05^2 + 3(2.05)(2.96) - 2.96^2] - [2^2 + 3(2)(3) - 3^2]\\ &= 0.6449.\end{aligned}$$

Therefore, $\Delta z \approx dz$.

Example 29 (Using differentials to estimate an error:)**.** The base radius and height of a right circular cone are measured as 10 cm and 25 cm, respectively, with a possible error in measurement of as much as 0.1 cm in each. Use differentials to estimate the maximum error in the calculated volume of the cone.

Solution.

Functions of Three or More Variables

Linear approximation:

$$\begin{aligned} f(x,y,z) &\approx L(x,y,z) \\ &= f(a,b,c) + f_x(a,b,c)(x-a) + f_y(a,b,c)(y-b) \\ &\quad + f_x(a,b,c)(z-c). \end{aligned}$$

Differentiability

Definition. Let $u = f(x,y,z)$. We say that f is differentiable at (a,b,c) if Δu can be expressed in the form

$$\begin{aligned} \Delta u = f_x(a,b,c)\Delta x + f_y(a,b,c)\Delta y + f_z(a,b,c)\Delta z + \varepsilon_1 \Delta x \\ + \varepsilon_2 \Delta y + \varepsilon_3 \Delta z, \end{aligned}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3 \to 0$ as $(\Delta x, \Delta y, \Delta z) \to (0,0,0)$.

Differentials

The differential du, also called the **total differential**, is defined by

$$du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy + \frac{\partial u}{\partial yz}dz.$$

Example 30. The dimensions of a rectangular box are measured to be 75 cm, 60 cm, and 40 cm, and each measurement is correct to within 0.2 cm. Use differentials to estimate the largest possible error when the volume of the box is calculated from these measurements.

Solution. If the dimensions of the box are x, y, and z, its volume is

$$v = xyz$$

and so

$$\begin{aligned} du &= \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy + \frac{\partial u}{\partial yz}dz \\ &= yzdx + xzdy + xydz. \end{aligned}$$

We are given that

$$\Delta x, \Delta y, \Delta z \leq 0.2.$$

To find the largest error in the volume, we therefore use

$$dx = dy = dz = 0.2$$

together with

$$x = 75, y = 60, z = 40.$$

We have

$$\Delta V = dV = (60)(40)(0.2)+(75)(40)(0.2)+(75)(60)(0.2) = 1980$$

Thus an error of only 0.2 cm in measuring each dimension could lead to an error of as much as 1980 cm in the calculated volume! This may seem like a large error, but it's only about 1% of the volume of the box.

6 Chain rule

Chain rule I: $z = f(x, y), x = x(t), y = y(t)$

Suppose that $x = x(t)$ and $y = y(t)$ are differentiable functions of t and $z = f(x, y)$ is a differentiable function of x and y. Then $z = f(x(t), y(t))$ is a differentiable function of t and

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

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y).

Proof.

A change of Δt in t produces changes of Δx in x and Δy in y . These, in turn, produce a change of Δz in z. We then have

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $(\Delta x, \Delta y) \to (0,0)$. Dividing both sides of this equation by Δt, we have

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \varepsilon_1 \frac{\Delta x}{\Delta t} + \varepsilon_2 \frac{\Delta y}{\Delta t},$$

If we now let $\Delta t \to 0$, then

$$\Delta x = x(t + \Delta t) - x(t) \to 0,$$

because g is differentiable therefore continuous. Similarly,

$$\Delta y \to 0.$$

This, in turn, means that $\varepsilon_1, \varepsilon_2 \to 0$, so

$$\begin{aligned}
\frac{dz}{dt} &= \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} \\
&= \frac{\partial f}{\partial x} \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y} \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} + \lim_{\Delta t \to 0} \varepsilon_1 \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \\
&\quad + \lim_{\Delta t \to 0} \varepsilon_2 \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} \\
&= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} + 0 \cdot \frac{dx}{dt} + 0 \cdot \frac{dy}{dt} \\
&= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.
\end{aligned}$$

◀

Example 31. If $z = x^2y + 3xy^4$, where $x = \sin 2t$ and $y = \cos t$, find dz/dt when $t = 0$.

The derivative in Example(31) can be interpreted as the rate of change of z with respect t to as the point (x, y) moves along the curve C with parametric equations $x = \sin 2t$, $y = \cos t$, (See the figure given below.)

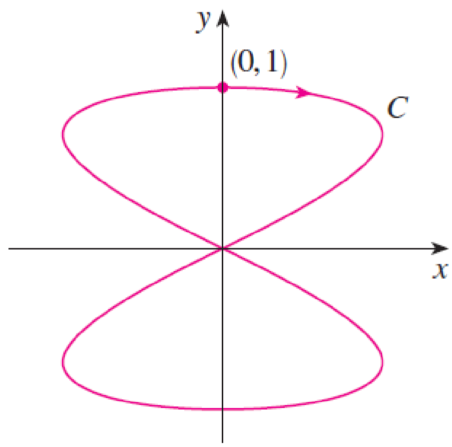


Figure 16: The curve $x = \sin 2t$, $y = \cos t$

In particular, when $t - 0$, the point (x, y) is (1,0) and $dz/dt = 6$ is the rate of increase as we move along the curve C through (1,0). If, for instance,

$$z = T(x, y) = x^2 + 3xy^4$$

represents the temperature at the point (x, y), then the composite function $z = T(\sin 2t, \cos t)$ represents the temperature at points on C and the derivative dz/dt represents the rate at which the temperature changes along C.

Chain rule II: $z = f(x, y), x = x(s, t), y = y(s, t)$

Suppose that $x = x(s, t)$ and $y = y(s, t)$ are differentiable functions of s, t and $z = f(x, y)$ is a differentiable function of x, y. Then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

Example 32. If $z = e^x \sin y$, where $x = st^2, y = s^2t$, find $\partial z/\partial s, \partial z/\partial t$.

It is easy to extend the chain rule to the general situation in which a dependent variable z is a function of n intermediate variables $x_1, ..., x_n$ each of which is, in turn, a function of m independent variables $t_1, ..., t_m$.

Example 33. Write out the Chain Rule for the case, where $w = f(x, y, z, t)$ and $x = x(u, v), y = y(u, v), z = z(u, v), t = t(u, v)$.

Example 34. If $g(s, t) = f(s^2 - t^2, t^2 - s^2)$ and f is differentiable, show that g satisfies the equation:

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 0.$$

Example 35. If $z = f(x, y)$ and $f$has continuous second-order partial derivatives and $x = r^2 + s^2, y = 2rs$, find $\partial z/\partial r, \partial^2 z/\partial r^2$.

Implicit Differentiation

(as an application of the chain rule.)

<u>Case I:</u> $F(x, y) = 0$, where $y = f(x)$.

Implicit Function Theorem:
If F is defined on a disk containing (a, b), where $F(a, b) = 0$, $F_y(a, b) \neq 0$, and F_x, F_y are continuous on the disk, then the equation $F(x, y) = 0$ defines y as a function of x near the point (a, b) and the derivative of this function is given by

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} = -\frac{F_x}{F_y}.$$

Example 36. Find y' if $x^3 + y^3 = 6xy$.

<u>Case II:</u> $F(x, y, z) = 0$, where $z = f(x, y)$.

Implicit Function Theorem:
If F is defined within a sphere containing (a, b, c), where $F(a, b, c) = 0$, $F_z(a, b, c) \neq 0$, and F_x, F_y, F_z are continuous inside the sphere, then the equation $F(x, y, z) = 0$ defines z as a function of x, y near the point (a, b, c) and the derivative of this function is differentiable, with partial derivatives given by

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

Example 37. Find $\partial z/\partial x$ and $\partial z/\partial y$ if $x^3 + y^3 + z^3 + 6xyz = 1$.

7 Directional derivatives and gradient vector

The partial derivatives of a function f tell us the rate of change of f in the direction of the coordinate axes. How can we measure the rate of change of f in other directions?

In order to formally define the derivative in a particular direction of motion, we want to represent the change in f for a given unit change in the direction of motion. We can represent this unit change in direction with a unit vector, say $u = (a, b)$. If we move a distance in the direction of u from a fixed point (x_0, y_0), we then arrive at the new point $(x_0 + ha, y_0 + hb)$. It now follows that the slope of the secant line to the curve on the surface through (x_0, y_0) in the direction of u through the points (x_0, y_0) and $(x_0 + ha, y_0 + hb)$ is

$$m_{\text{sec}} = \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}. \tag{1}$$

To get the instantaneous rate of change of f in the direction $u = (a, b)$, we must take the limit of the quantity in Equation (1) as $h \to 0$. Doing so results in the formal definition of the directional derivative.

Definition:
The directional derivative of f at (x_0, y_0) in the direction of a unit vector $u = (a, b)$ is

$$D_u f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if this limit exists.

By comparing this definition with the definitions of partial derivatives:

$$f_x(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$
$$f_y(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h},$$

we see that if $u = i = (1, 0)$, then $D_i f = f_x$ and if $u = j = (0, 1)$, then $D_j f = f_y$. In other words, the partial derivatives of f with respect to x and y are just special cases of the directional derivative.

It is time consuming to find the directional derivative using the above definition. However, we can find a way to evaluate directional derivatives without resorting to the limit definition.

Theorem 7.1. *If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $u = (a, b)$ and*

$$D_u f(x, y) = f_x(x, y)a + f_y(x, y)b. \tag{8}$$

Proof.

If we define a function of the single variable h by

$$g(h) = f(x_0 + ha, y_0 + hb)$$

then, by the definition of a derivative, we have

$$\begin{aligned} g'(0) &= \lim_{h \to 0} \frac{g(h) - g(0)}{h} \\ &= \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h} \\ &= D_u f(x_0, y_0). \end{aligned}$$

On the other hand, we can write $g(h) = f(x, y)$, where $x = x_0 + ha, y = y_0 + hb$, so the Chain Rule gives

$$\begin{aligned} g'(h) &= \frac{\partial f}{\partial x}\frac{dx}{dh} + \frac{\partial f}{\partial y}\frac{dy}{dh} \\ &= f_x(x, y)a + f_y(x, y)b. \end{aligned}$$

If we now put $h = 0$, then $x = x_0, y = y_0$, and

$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

Therefore, we see that

$$D_u f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

◀

Remark 1.

To use the theroem, we must have a unit vector in the direction of motion. In the event that we have a direction prescribed by a non-unit vector, we must first scale the vector to have length 1.

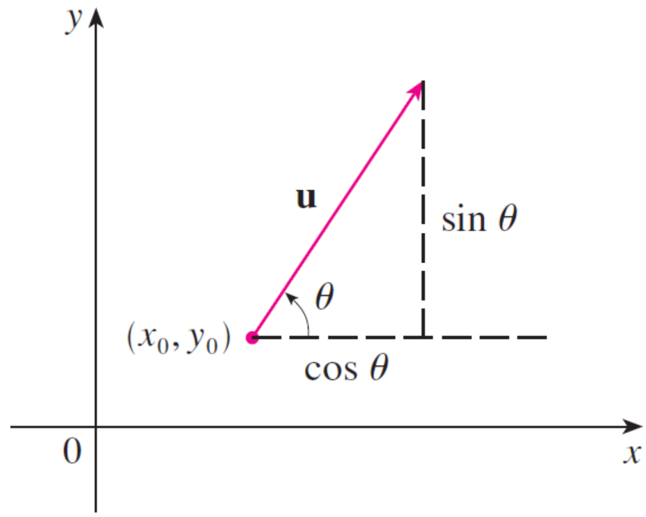


Figure 17: A unit vector $\vec{u} = (a, b) = (\cos\theta, \sin\theta)$

If the unit vector u makes an angle θ with the positive x-axis (as in Figure 17), then we can write $u = (\cos\theta, \sin\theta)$ and Formula (8) becomes

$$D_u f(x, y) = f_x(x, y)\cos\theta + f_y(x, y)\sin\theta.$$

Example 38. Find the directional derivative $D_u f(x, y)$ if

$$f(x, y) = x^3 - 3xy + 4y^2$$

and u is the unit vector given by angle $\theta = \pi/6$. What is $D_u f(1, 2)$?

The Gradient Vector

Notice that the directional derivative of a differentiable function can be written as the dot product of two vectors:

$$\begin{aligned} D_u f(x,y) &= f_x(x,y)a + f_y(x,y)b \\ &= (f_x(x,y), f_y(x,y)) \cdot (a,b) \\ &= (f_x(x,y), f_y(x,y)) \cdot u. \end{aligned}$$

The first vector in this dot product is called the **gradient** of f and is denoted by

$$\operatorname{grad} f \quad \text{or} \quad \nabla f.$$

Definition:
If f is a function of two variables x and y, then the gradient of f is the vector function ∇f defined by

$$\nabla f(x,y) = (f_x(x,y), f_y(x,y)) = f_x i + f_y j.$$

Example 39.

Find the gradient f if

$$f(x,y) = \sin x + e^{xy}.$$

What is $\nabla f(0,1)$?

With this notation for the gradient vector, we can rewrite the expression (8) for the directional derivative of a differentiable function as

$$\boxed{D_u f(x,y) = \nabla f(x,y) \cdot u.}$$

This expresses the directional derivative in the direction of u as the scalar projection of the gradient vector onto u.

Example 40.

> **Using a gradient vector to find a directional derivative:**
> Find the directional derivative of the function $f(x, y) = x^2y^3 - 4y$ at the point $(2, -1)$ in the direction of the vector $v = 2i + 5j$.

Functions of Three Variables

For functions of three variables we can define directional derivatives in a similar manner. Again $D_uf(x, y, z)$ can be interpreted as the rate of change of the function in the direction of a unit vector $\vec{u}$.

> **Definition:**
> The directional derivative of f at $\vec{x}_0 = (x_0, y_0, z_0)$ in the direction of a unit vector $\vec{u} = (a, b, c)$ is
>
> $$D_uf(\vec{x}_0) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{u}) - f(\vec{x}_0)}{h}$$
>
> if this limit exists.

This is reasonable because the vector equation of the line through $\vec{x}_0$ in the direction of the vector $\vec{u}$ is given by

$$\vec{x} + t\vec{u}$$

and so $f(\vec{x} + h\vec{u})$ represents the value of f at a point on this line.

If $f(x, y, z)$ is differentiable and $\vec{u} = (a, b, c)$, then we can prove by the same method as in the case of a function of two variables that

$$D_{\vec{u}} f(x, y, z) = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c. \tag{9}$$

For a function f of three variables, the gradient vector is

$$\nabla f(x, y, z) = (f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)).$$

or

$$\nabla f = (f_x, f_y, f_z) = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k$$

Then, just as with functions of two variables, Formula (9) for the directional derivative can be rewritten as

$$D_{\vec{u}} f(x, y, z) = \nabla f_x(x, y, z) \cdot u.$$

Example 41.

If $f(x, y, z) = x \sin yz$,

(a) find the gradient of f

(b) find the directional derivative of f at (1, 3, 0) in the direction of $v = i + 2j - k$.

Maximizing the Directional Derivative

Theorem 7.2. *Suppose f is a differentiable function of two or three variables. Then the maximum value of the directional derivative $D_{\vec{u}f(x)}$ is $|\nabla f(x)|$ and it occurs when $\vec{u}$ has the same direction as the gradient vector $\nabla f(x)$.*

Proof.

We have

$$\begin{aligned} D_{\vec{u}} f &= \nabla f \cdot \vec{u} \\ &= |\nabla f||\vec{u}| \cos\theta \\ &= |\nabla f| \cos\theta, \end{aligned}$$

where θ is the angle between ∇f and $\vec{u}$. The maximum value of $\cos\theta$ is 1 and this occurs when $\theta = 0$. Therefore the maximum value of $D_{\vec{u}} f$ is $|\nabla f|$ and it occurs when $\theta = 0$, that is, when $\vec{u}$ has the same direction as ∇f. ◀

Example 42.

Let $f(x, y, z) = xe^y$.

(a) find the directional derivative of f at $P(2, 0)$ in the direction from P to $Q(1/2, 2)$.

(b) In what direction does f have the maximum rate of change? What is this maximum rate of change?

Example 43.

Suppose that the temperature at a point (x, y, z) in space is given by

$$T(x, y, z) = 80/(1 + x^2 + 2y^2 + 3z^2),$$

where T is measured in degrees Celsius and x, y, z in meters. In which direction does the temperature increase fastest at the point $(1, 1, -2)$? What is the maximum rate of increase?

Tangent Planes to Level Surfaces

Suppose S is a surface with equation $F(x, y, z) = k$, that is, it is a level surface of a function F of three variables, and let $P(x_0, y_0, z_0)$ be a point on S. Let C be any curve that lies on the surface S and passes through the point P. The curve C is described by a continuous vector function $\vec{r}(t) = (x(t), y(t), z(t))$. Let t_0 be the parameter value corresponding to P; that is, $\vec{r}(t_0) = (x_0, y_0, z_0)$. Since C lies on s, any point $(x(t), y(t), z(t))$ must satisfy the equation of s, that is,

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

If x, y, z are differentiable functions of t and F is also differentiable, then we can use the Chain Rule to differentiate its both sides to obtain

$$\nabla F \cdot \vec{r}'(t) = 0.$$

In particular, when $t = t_0$ we have $\vec{r}(t_0) = (x_0, y_0, z_0)$, so

$$\nabla F(x_0, y_0, z_0) \cdot \vec{r}'(t_0) = 0.$$

This equation says that the gradient vector $\nabla F(x_0, y_0, z_0)$ at P is perpendicular to the tangent vector $\vec{r}'(t_0)$ to any curve C on S that passes through P. (See the figure given below.)

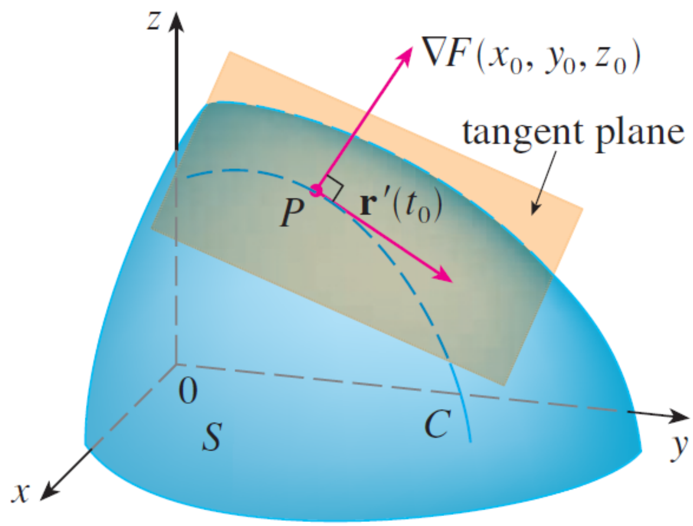


Figure 18

If $\nabla F(x_0, y_0, z_0) \neq 0$, it is therefore natural to define the **tangent plane to the level surface** $F(x_0, y_0, z_0) = k$ at $P(x_0, y_0, z_0)$ as the plane that passes through P and has normal vector $\nabla F(x_0, y_0, z_0)$. Using the standard equation of a plane, we can write the equation of this tangent plane as

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) \\ + F_z(x_0, y_0, z_0)(z - z_0) = 0. \qquad (10)$$

The normal line to S at P is the line passing through P and perpendicular to the tangent plane. The direction of the normal line is therefore given by the gradient vector $\nabla F(x_0, y_0, z_0)$ and so, its symmetric equations are

$$\boxed{\frac{x-x_0}{F_x(x_0,y_0,z_0)} = \frac{y-y_0}{F_y(x_0,y_0,z_0)} = \frac{z-z_0}{F_z(x_0,y_0,z_0)}}$$

In the special case in which the equation of a surface S is of the form $z = f(x, y, z)$ (that is, is the graph of a function f of two variables), we can rewrite the equation as

$$F(x, y, z) = f(x, y, z) - z = 0$$

and regard S as a level surface (with $k = 0$) of F. Then

$$\begin{aligned} F_x(x_0, y_0, z_0) &= f_x(x_0, y_0) \\ F_y(x_0, y_0, z_0) &= f_y(x_0, y_0) \\ F_z(x_0, y_0, z_0) &= -1 \end{aligned}$$

so Equation (10) becomes

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

Thus our new, more general, definition of a tangent plane is consistent with the definition that was given earlier.

Example 44.

Find the equations of the tangent plane and normal line at the point $(-2, 1, -3)$ to the ellipsoid

$$\frac{x^2}{4} + y^2 + \frac{z^2}{9} = 3.$$

8 Maximum and minimum values

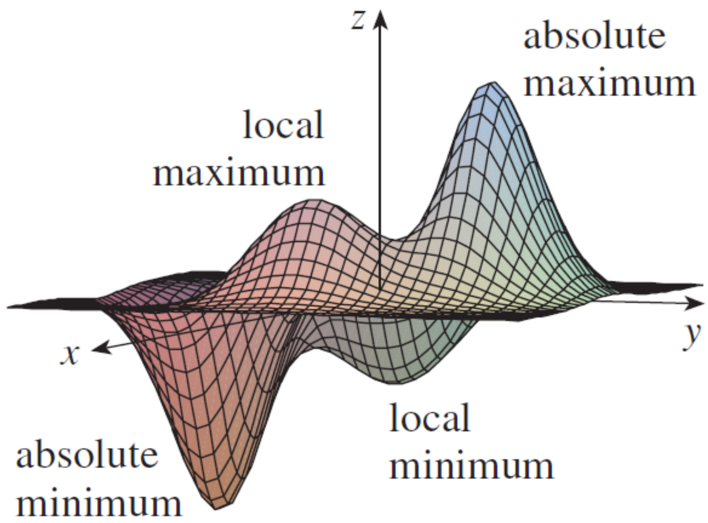


Figure 19

Definition. Let f be a function of two variables x and y.

- The function f has a **local maximum** at a point (a, b) provided that $f(x, y) \leq f(a, b)$ for all points (x, y) in some disk with center at (a, b). In this situation we say that $f(a, b)$ is a **local maximum value**.
- The function f has a **local minimum** at a point (a, b) provided that $f(x, y) \geq f(a, b)$ for all points (x, y) in some disk with center at (a, b). In this situation we say that $f(a, b)$ is a **local minimum value**.
- An **absolute maximum point** is a point (a, b) for which $f(x, y) \leq f(a, b)$ for all points (x, y) in the domain of f. The value of f at an absolute maximum point is the **maximum value** of f.
- An **absolute minimum point** is a point such that $f(x, y) \geq f(a, b)$ for all points (x, y) in the domain of

f. The value of f at an absolute minimum point is the **maximum value** of f.

We use the term **extremum point** to refer to any point (a, b) at which f has a local maximum or minimum. In addition, the function value $f(a, b)$ at an extremum is called an **extremal value**.

Definition. A **critical point** (a, b) of a function $f = f(x, y)$ is a point in the domain of f at which $f_x(a, b) = 0$ and $f_y(a, b) = 0$, or such that one of $f_x(a, b)$ or $f_y(a, b)$ fails to exist.

We can therefore find critical points of a function f by computing partial derivatives and identifying any values of (x, y) for which one of the partials doesnt exist or for which both partial derivatives are simultaneously zero.

Definition. A **stationary** (a, b) of a function $f = f(x, y)$ is a point in the domain of f at which $f_x(a, b) = f_y(a, b) = 0$.

Theorem 8.1 (Fermat)**.** *If f has a local maximum or minimum at (a, b) and the first-order partial derivatives of f exist there, then $f_x(a, b) = 0$ and $f_y(a, b) = 0$.*

Proof.

Let $g(x) = f(x, b)$. If f has a local maximum (or minimum) at (a, b), then g has a local maximum (or minimum) at a, so by Fermat's Theorem for functions of one variable. But $g'(a) = f_x(a, b)$ and so

$f_x(a, b) = 0$. Similarly, by applying Fermat's Theorem to the function $G(y) = f(a, y)$, we obtain $f_y(a, b) = 0$. ◀

If we put $f_x(a, b) = 0$ and $f_y(a, b) = 0$ in the equation of a tangent plane:

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

we get

$$z = z_0.$$

Thus the geometric interpretation of Fermat's theorem is that

> If the graph of f has a tangent plane at a local maximum or minimum, then the tangent plane must be horizontal.

This theorem says that if f has a local maximum or minimum at (a, b), then (a, b) is a critical point of f. All critical points give rise to maxima or minima. At a critical point, a function could have a local maximum or a local minimum or neither.

Example 45.

Investigate the critical points of the function:

$$f(x, y) = x^2 + y^2 - 2x - 6y + 14.$$

Solution. We have

$$f_y(x, y) = 2x - 2 \qquad f_y(x, y) = 2y - 6.$$

These partial derivatives are equal to 0 when $x = 1$ and y=3, so the only critical point is (1,3). By completing the square, we find that

$$f(x, y) = 4 + (x - 1)^2 + (y - 3)^2.$$

Since $(x - 1)^2, (y - 3)^2 \geq 0$, we have $f(x, y) \geq 4$ for all values of x, y. Therefore $f(1, 3) = 4$ is a local minimum, and in fact it is the absolute minimum of f. This can be confirmed geometrically from the graph of f which is the elliptic paraboloid with vertex shown in the figure geven below.

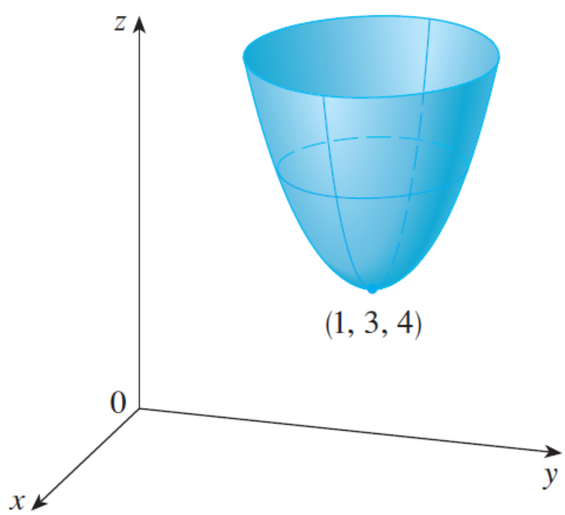


Figure 20: The paraboloid $f(x, y) = x^2 + y^2 - 2x - 6y + 14$

Example 46.

A function with no extreme values: Investigate the extreme values of $f(x, y) = xy$.

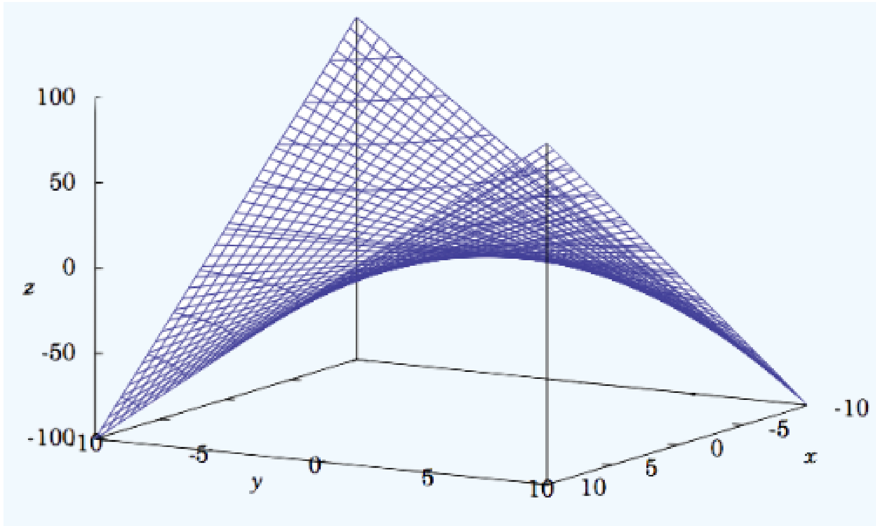


Figure 21: The hyperbolic paraboloid $f(x, y) = x^2 + y^2 - 2x - 6y + 14$

This example illustrates the fact that a function need not have a maximum or minimum value at a critical point. Figure 3 shows how this is possible. The graph f of is the hyperbolic paraboloid $z = xy$, which has a horizontal tangent plane ($z = 0$) at the origin. You can see that $f(0, 0) = 0$ is a maximum in the direction of the line $y = x$ but a minimum in the direction of the line $y = -x$. Near the origin the graph has the shape of a saddle and so the point $(0, 0)$is called a *saddle point* of f.

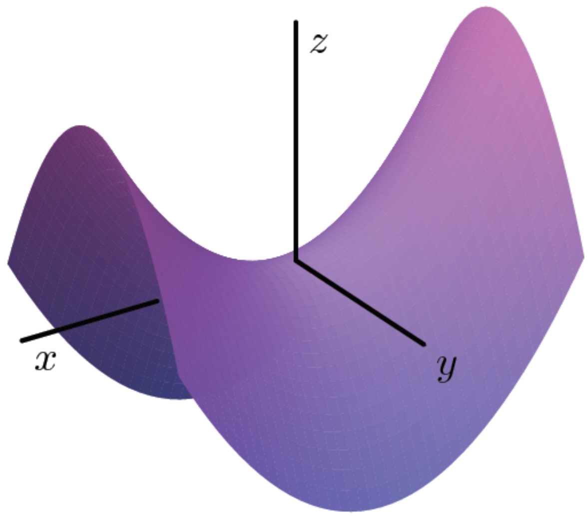


Figure 22: (0,0): saddle point

Definition.
Given the function $z = f(x, y)$, the point $(a, b, f(a, b))$ is a **saddle point** if both $f_x(a, b) = 0$ and $f_y(a, b) = 0$, but f does not have a local extremum at (a, b).

The Second Derivative Test.

Suppose the second partial derivatives of f are continuous on a disk with center (a, b) and suppose that $f_x(a, b) = 0$ and $f_y(a, b) = 0$. Let D be the quantity defined by

$$D = f_{xx}(a, b) f_{yy}(a, b) - f_{xy}(a, b)^2.$$

1. If $D > 0$ and $f_{xx}(a, b) < 0$, then f has a local maximum at (a, b).
2. If $D > 0$ and $f_{xx}(a, b) > 0$, then f has a local minimum at (a, b).
3. If $D < 0$, then f has a saddle point at (a, b).
4. If $D = 0$, then this test yields no information about what happens at (a, b).

The quantity D is called the **discriminant** of the function f at (a, b).

Remark 2.

It's helpful to write D as a determinant:

$$D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx} f_{yy} - (f_{xy})^2.$$

The matrix

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

is called the **Hessian matrix** of f.

Example 47.

Find the local maximum and minimum values and saddle points of

$$f(x, y) = x^4 + y^4 - 4xy + 1.$$

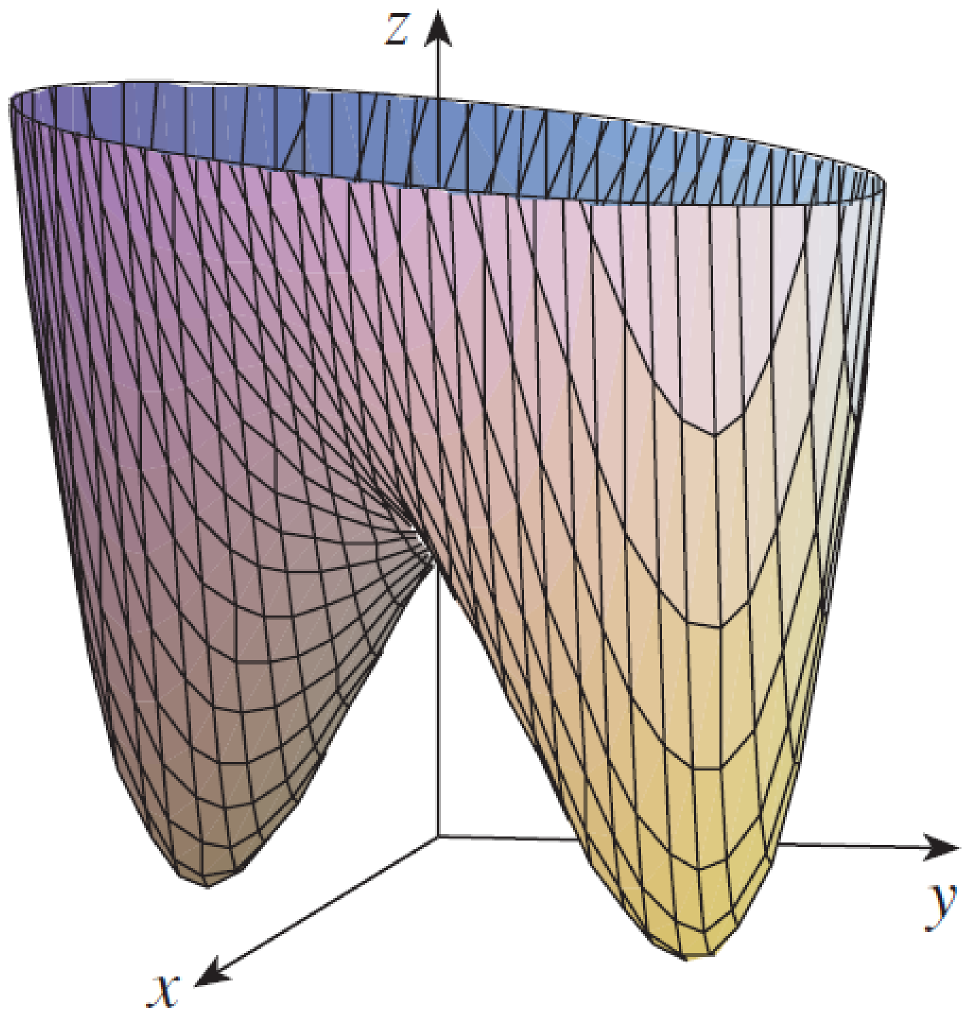


Figure 23: $f(x, y) = x^4 + y^4 - 4xy + 1$

Example 48.

Find all local maxima and minima of

$$f(x, y) = (x^2 + y^2)e^{(x^2+y^2)}.$$

Solution. First find the critical points, i.e. where $\nabla f = 0$. Since

$$\frac{\partial f}{\partial x} = 2x(1-(x^2+y^2))e^{-(x^2+y^2)}$$
$$\frac{\partial f}{\partial y} = 2y(1-(x^2+y^2))e^{-(x^2+y^2)},$$

then the critical points are $(0,0)$ and all points (x,y) on the unit circle $x^2+y^2=1$.

Now, the second-order partial derivatives are:

$$\frac{\partial^2 f}{\partial x^2} = 2[1-(x^2+y^2)-2x^2-2x^2(1-(x^2+y^2))]e^{-(x^2+y^2)}$$
$$\frac{\partial^2 f}{\partial y^2} = 2[1-(x^2+y^2)-2y^2-2y^2(1-(x^2+y^2))]e^{-(x^2+y^2)}$$
$$\frac{\partial^2 f}{\partial y \partial x} = -4xy[2-(x^2+y^2)]e^{-(x^2+y^2)}$$

At $(0,0)$, we have $D = 4 > 0$ and $\frac{\partial^2 f}{\partial x^2}(0,0) = 2 > 0$, so $(0,0)$ is a local minimum. However, for points (x,y) on the unit circle $x^2+y^2=1$, we have

$$D = (-4x^2e^{-1})(-4y^2e^{-1}) - (-4xye^{-1})^2 = 0$$

and so the test fails.

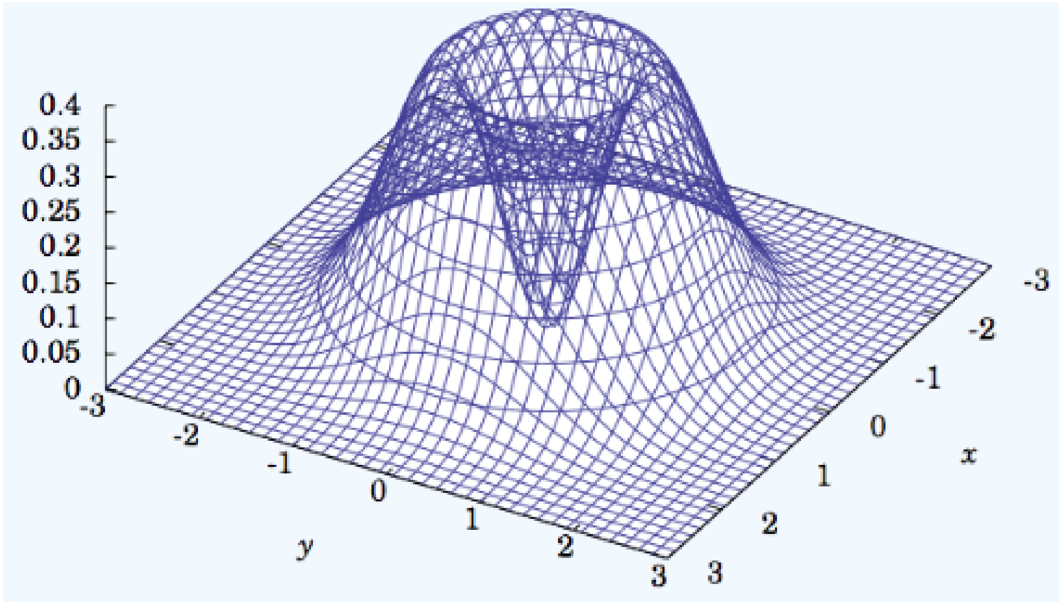


Figure 24: $f(x, y) = (x^2 + y^2)e^{(x^2+y^2)}$

If we look at the graph of $f(x, y)$, as shown in the above figure, it looks like we might have a local maximum for (x, y) on the unit circle $x^2 + y^2 = 1$. If we switch to using polar coordinates (r, θ) instead of (x, y) in $\mathbb{R}^2$, where $r^2 = x^2 + y^2$, then we see that we can write $f(x, y)$ as a function $g(r)$ of the variable r alone:

$$g(r) = r^2 e^{-r^2}.$$

Then

$$g'(r) = r^2(1 - r^2)e^{-r^2},$$

so it has a critical point at $r = 1$, and we can check that

$$g"(1) = -4e^{-1} < 0,$$

so the Second Derivative Test from single-variable calculus says that $r = 1$ is a local maximum. But $r = 1$ corresponds to the unit circle $x^2 + y^2 = 1$. Thus, the points (x, y) on the unit circle $x^2 + y^2 = 1$ are local maximum points for f .

Example 49.

Find and classify all the critical points of $f(x, y) = 4 + x^3 + y^3 - 3xy$.

Solution. We first need all the first order (to find the critical points) and second order (to classify the critical points) partial derivatives so let's get those.

$f_x = 3x^2 - 3y$ $\quad$ $f_y = 3y^2 - 3x$

$f_{xx} = 6x$ $\quad$ $f_{yy} = 6y$ $\quad$ $f_{xy} = -3$

$$\begin{aligned} D(x, y) &= f_{xx}(x, y)f_{yy}(x, y) - f_{xy}(x, y)^2 \\ &= 6x \times 6y - (-3)^2 \\ &= 36xy - 9. \end{aligned}$$

Let's first find the critical points. Critical points will be solutions to the system of equations:

$$\begin{aligned} f_x &= 3x^2 - 3y = 0 \\ f_y &= 3y^2 - 3x = 0. \end{aligned}$$

This is a non-linear system of equations and these can, on occasion, be difficult to solve. However, in this case it's not too bad. We can solve the first equation for y as follows:

$$3x^2 - 3y = 0 \Rightarrow y = x^2.$$

Plugging this into the second equation gives,

$$3x^4 - 3x = 3x(x^3 - 1) = 0$$

From this we can see that we must have $x = 0$ or $x = 1$. Now use the fact that $y = x^2$ to get the critical points.

$$x = 0 \Rightarrow y = 0$$
$$x = 1 \Rightarrow y = 1.$$

So, we get two critical points:

$$(0, 0), (1, 1).$$

All we need to do now is classify them. To do this we will need the sign of D at critical points. We have

$$D(0, 0) = -9 < 0.$$

So, the critical point (0,0) must be a saddle point.

We also have

$$D(1, 1) = 36(1)(1) - 9 = 27 > 0$$

and

$$f_{xx}(1, 1) = 6(1) = 6 > 0.$$

Therefore, f has a local minimum at (1,1).

Thus,

Saddle point at $(0, 0)$,

Relative minimum at $(1, 1)$.

Example 50.

Find and classify all the critical points for $f(x, y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$.

Example 51.

Find the shortest distance from the point $(1, 0, -2)$ to the plane $x + 2y + z = 4$.

Solution. Here, we need to first come up with the equation that we are going to have to work with.

First, lets suppose that (x, y, z) is any point on the plane. The distance between this point and the point in question, $(1, 0, -2)$, is given by the formula,

$$d = \sqrt{(x-1)^2 + y^2 + (z+2)^2}.$$

What we are then asked to find is the minimum value of d. The point (x, y, z) that gives the minimum value of d will be the point on the plane that is closest to $(1, 0, -2)$.

There are a couple of issues with this equation. First, it is a function of x, y and z and we can only deal with functions of x and y at this point. However, this is easy to fix. We can solve the equation of the plane to see that,

$$z = 4 - x - 2y.$$

Plugging this into the distance formula gives

$$\begin{aligned} d &= \sqrt{(x-1)^2 + y^2 + (4 - x - 2y + 2)^2} \\ &= \sqrt{(x-1)^2 + y^2 + (6 - x - 2y)^2} \end{aligned}$$

Now, the next issue is that there is a square root in this formula and we know that we're going to be differentiating this eventually. So, in order to make our life a little easier let's notice that finding the minimum value of d will be equivalent to finding the minimum value of d^2.

So, let's instead find the minimum value of

$$d^2 = (x-1)^2 + y^2 + (6 - x - 2y)^2.$$

We can minimize d by minimizing the simpler expression

$$f(x, y) = (x - 1)^2 + y^2 + (6 - x - 2y)^2.$$

Now, we need to be a little careful here. We are being asked to find the closest point on the plane to $(1, 0, -2)$ and that is not really the same thing as finding and classifying critical points as relative minimums or maximums. What we are really asking is to find the smallest value (or the absolute minimum) of the function d. However, we know that the equation

$$ax^2 + by^2 + cz^2 = d$$

represents an ellipsoid. Hence in this case, a relative minimum is an absolute minimum as well.

So, lets go through the process. We have ...

Example 52.

A rectangular box without a lid is to be made from 12 m of cardboard. Find the maximum volume of such a box.

Solution. Let the length, width, and height of the box (in meters) be x, y, and z. Then the volume of the box is

$$V = xyz.$$

We can express V as a function of just two variables and by using the fact that the area of the four sides and the bottom of the box is

$$2xz + 2yz + xy = 12 \Rightarrow z = \frac{12 - xy}{2(x + y)}.$$

Thus,

$$V = xy\frac{12 - xy}{2(x+y)} = \frac{12xy - x^2y^2}{2(x+y)}.$$

Now, lets go through the process. We have ...

Example 53.

Find the local extrema of

$$f(x, y) = x^3 + x^2y - y^2 - 4y.$$

Solution. Step 1: Find the critical points. The derivative of f is

$$Df(x, y) = \begin{bmatrix} 3x^2 + 2xy & x^2 - 2y - 4 \end{bmatrix}.$$

$Df(x, y) = [0 \quad 0]$ means both components must be zero simultaneously. We need

$$x(3x + 2y) = 0 \qquad (1)$$

and

$$x^2 - 2y - 4 = 0. \qquad (2)$$

We need to solve two equations for the two unknowns x and y.

Equation (1) is satisfied if either $x = 0$ or if $3x + 2y = 0$, i.e., if $x = 0$ or if $y = -3x/2$. We consider these two solutions as two separate cases. For each case, we will find soultions for equation (2).

Case 1: Let $x = 0$. Then we know equation (1) is satisfied. We plug $x = 0$ into equation (2), which becomes $0 - 2y - 4 = 0$, i.e., $y = 2$. If $x = 0$ and $y = -2$, then both equation (1) and equation (2) are satisfied. Therefore the point $(0, -2)$ is a critical point.

Case 2: Let $y = -3x/2$. Then we know that Equation (1) is satisfied. We plug $y = -3x/2$ into Equation (2) and simplify:

$$\begin{aligned} & x^2 - 2(-3x/2) - 4 = 0 \\ \Rightarrow\ & x^2 + 3x - 4 = 0 \\ \Rightarrow\ & (x-1)(x+4) = 0 \\ \Rightarrow\ & x = 1 \text{ or } x = -4. \end{aligned}$$

So, we have two solutions of equation (2) for case 2. The first solution is when $x = 1$, which means $y = -3x/2 = -3/2$. If $x = 1$ and $y = -3/2$, then both equation (1) and equation (2) are satisfied. Therefore the point $(1, -3/2)$ is a critical point.

The second solution for case 2 is when $x = -4$, which means $y = -3x/2 = 6$. Therefore, the point $(-4, 6)$ is a critical point.

To summarize the results from both case 1 and case 2, we conclude that $f(x, y)$ has three critical points: $(0, -2)$, $(1, -3/2)$, and $(-4, 6)$.

Step 2: Classify the critical points.

The Hessian matrix is

$$Hf(x, y) = \begin{bmatrix} 6x + 2y & 2x \\ 2x & -2 \end{bmatrix}$$

We need to check the definiteness of the $Hf(x, y)$ at the critical points $(0, -2), (1, -3/2)$, and $(-4, 6)$.

For the critical point $(0,-2)$,

$$Hf(0,-2) = \begin{bmatrix} -4 & 0 \\ 0 & -2 \end{bmatrix}$$

$h_{11} = -4 < 0$ and $\det(Hf) = 8 > 0$. This means $Hf(0,-2)$ is negative definite and f has a local maximum at $(0,-2)$.

For the critical point $(1,-3/2)$,

$$Hf(-4,6) = \begin{bmatrix} -12 & -8 \\ -8 & -2 \end{bmatrix}.$$

$h_{11} = 3 > 0$ and $\det(Hf) = -6 - 4 = -10 < 0$. This means $Hf(1,-3/2)$ is indefinite and f has a saddle at $(1,-3/2)$.

For the critical point $(-4,6)$,

$$Hf(-4,6) = \begin{bmatrix} -12 & -8 \\ -8 & -2 \end{bmatrix}.$$

$h_{11} = -12 < 0$ and $\det(Hf) = 24 - 64 = -40 < 0$. This means $Hf(-4,6)$ is indefinite and f has a saddle at $(-4,6)$.

Example 2 Identify the local extrama of $f(x,y) = (x^2+y^2)e^{-y}$.

Solution

Step 1: Find the critical points.

The derivative of f is

$$Df(x,y) = \begin{bmatrix} 2xe^{-y} & (2y - x^2 - y^2)e^{-y} \end{bmatrix}$$

$Df(x, y) = [0 \quad 0]$ means that $2x = 0$ and $2yx^2y^2 = 0$, i.e., $x = 0$ and $y(2y) = 0$.

The critical points are therefore $(0, 0)$ and $(0, 2)$.

Step 2: Classify the critical points.

The Hessian matrix is

$$Hf(x, y) = \begin{bmatrix} 2e^{-y} & -2xe^{-y} \\ -2xe^{-y} & (2 - 4y + y^2 + x^2)e^{-y} \end{bmatrix}$$

At the critical point $(0, 0)$

$$Hf(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$h_{11} = 2 > 0$ and $\det(Hf) = 4 > 0$, so $(0, 0)$ is a local mininum.

At the critical point $(0, 2)$

$$Hf(0, 2) = \begin{bmatrix} e^{-2} & 0 \\ 0 & -2e^{-2} \end{bmatrix}$$

$h_{11} = e^2 > 0$ and $\det(Hf) = 2e^4 < 0$ so $(0, 2)$ is a saddle point.

9 Lagrange multipliers

Method of Lagrange Multipliers To find the maximum and minimum values of $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$ [assuming that these extreme values exist and $\nabla g \neq 0$ on the surface $g(x, y, z) = k$]:

(a) Find all values of x, y, z and λ solving the system of equations:

$$\begin{aligned} &\nabla f(x, y, z) = \lambda \nabla g(x, y, z) \\ &g(x, y, z) = k. \end{aligned}$$

(b) Evaluate f at all the points (x, y, z) that result from step (a). The largest of these values is the maximum value of f, the smallest is the minimum value of f.

Notice that

- The system of equations from the method actually has four equations. To see this let's take the first equation and put in the definition of the gradient vector to see what we get.

  $$(f_x, f_y, f_z) = \lambda(g_x, g_y, g_z) = (\lambda g_x, \lambda g_y, \lambda g_z)$$

  In order for these two vectors to be equal the individual components must also be equal. So, we actually have three equations here.

  $$f_x = \lambda g_x \quad f_y = \lambda g_y \quad f_z = \lambda g_z$$

  These three equations along with the constraint

  $$g\,(x, y, z) = c,$$

give four equations with four unknowns x, y, z, and λ.

- If we only have functions of two variables then we won't have the third component of the gradient and so will only have three equations in three unknowns x, y, and λ.

Proof. Suppose that (a, b, c) is a point of S and that $f(x, y, z) \geq f(a, b, c)$ for all points (x, y, z) on S that are close to (a, b, c). That is (a, b, c) is a local minimum for f on S. Of course, the argument for a local maximum is virtually identical.
Imagine that we go for a walk on S, with the time t running, say, from $t = -1$ to $t = +1$ and that at time $t = 0$ we happen to be exactly at (a, b, c). Let's say that our position is $\big(x(t), y(t), z(t)\big)$ at time t. (We are always on S, so $g\big(x(t), y(t), z(t)\big) = 0$ for all t.)
Write

$$F(t) = f\big(x(t), y(t), z(t)\big)$$

So $F(t)$ is the value of f that we see on our walk at time t. Then for all t close to 0, the point $\big(x(t), y(t), z(t)\big)$ is close to

$$\big(x(0), y(0), z(0)\big) = (a, b, c)$$

so that

$$\begin{aligned} F(0) &= f\big(x(0), y(0), z(0)\big) = f(a, b, c) \\ &\leq f\big(x(t), y(t), z(t)\big) = F(t) \end{aligned}$$

for all t close to zero. So $F(t)$ has a local minimum at $t = 0$ and consequently $F'(0) = 0$.

By the chain rule,

$$\begin{aligned}
F'(0) &= \frac{d}{dt} f\big(x(t), y(t), z(t)\big)\Big|_{t=0} \\
&= f_x\big(a,b,c\big)x'(0) + f_y\big(a,b,c\big)y'(0) + f_z\big(a,b,c\big)z'(0) \\
&= 0
\end{aligned}$$

We may rewrite this as a dot product:

$$\begin{aligned}
0 = F'(0) &= \nabla f(a,b,c) \cdot (x'(0)\,,\ y'(0)\,,\ z'(0)) \\
\implies \nabla f(a,b,c) &\perp (x'(0)\,,\ y'(0)\,,\ z'(0))
\end{aligned}$$

This is true for all paths on S that pass through (a,b,c) at time 0. In particular it is true for all vectors $(x'(0)\,,\ y'(0)\,,\ z'(0))$ that are tangent to S at (a,b,c). So $\nabla f(a,b,c)$ is perpendicular to S at (a,b,c).
But we already know that $\nabla g(a,b,c)$ is also perpendicular to S at (a,b,c). So $\nabla f(a,b,c)$ and $\nabla g(a,b,c)$ have to be parallel vectors. That is,

$$\nabla f(a,b,c) = \lambda \nabla g(a,b,c)$$

for some number λ. That's the Lagrange multiplier rule of our theorem. ◀

Example 54.

> Find the maximum and minimum of the function $x^2 - 10x - y^2$ on the ellipse whose equation is $x^2 + 4y^2 = 16$.

Solution. For this problem the objective function is $f(x,y) = x^2 - 10x - y^2$ and the constraint function is $g(x,y) = x^2 + 4y^2 - 16$. To apply the method of Lagrange multipliers we need ∇f and ∇g. So we start by computing the first order derivatives of these functions.

$$f_x = 2x - 10 \qquad f_y = -2y \qquad g_x = 2x \qquad g_y = 8y$$

So, according to the method of Lagrange multipliers, we need to find all solutions to

$$\begin{aligned} 2x - 10 &= \lambda(2x) \\ -2y &= \lambda(8y) \\ x^2 + 4y^2 - 16 &= 0 \end{aligned} \tag{$*$}$$

Rearranging these equations gives

$$(\lambda - 1)x = -5 \tag{E1}$$
$$(4\lambda + 1)y = 0 \tag{E2}$$
$$x^2 + 4y^2 - 16 = 0 \tag{E3}$$

From (E2), we see that we must have either $\lambda = -\frac{1}{4}$ or $y = 0$.

- If $\lambda = -\frac{1}{4}$, (E1) gives $-\frac{5}{4}x = -5$, i.e. $x = 4$, and then (E3) gives $y = 0$.
- If $y = 0$, then $(*)$ gives $x = \pm 4$.

So the method of Lagrange multipliers gives that the only possible locations of the maximum and minimum of the function f are $(4, 0)$ and $(-4, 0)$. To complete the problem, we only have to compute f at those points.

point	$(4,0)$	$(-4,0)$
value of f	-24	56
	min	max

Hence the maximum value of $x^2 - 10x - y^2$ on the ellipse is 56 and the minimum value is -24.

In the previous example, the objective function and the constraint were specified explicitly. That will not always be the case. In the next example, we have to do a little geometry to extract them.

Example 55.

Find the rectangle of largest area (with sides parallel to the coordinates axes) that can be inscribed in the ellipse $x^2 + 2y^2 = 1$.

Solution. Since this question is so geometric, it is best to start by drawing a picture.

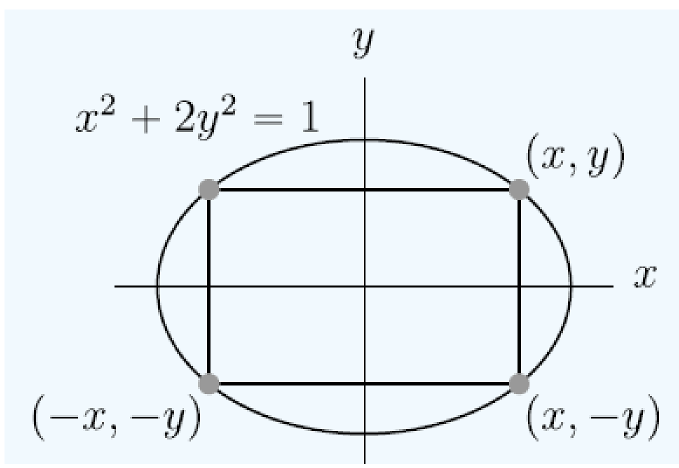


Figure 25

Call the coordinates of the upper right corner of the rectangle (x, y), as in the figure above. The four corners of the rectangle are $(\pm x, \pm y)$ so the rectangle has width $2x$ and height $2y$ and we have the problem:

$$\text{Maximize } f(x,y) = 4xy$$
$$\text{subject to } g(x,y) = x^2 + 2y^2 - 1.$$

Again, to use Lagrange multipliers we need the first order partial derivatives.

$$f_x = 4y \qquad f_y = 4x \qquad g_x = 2x \qquad g_y = 4y$$

So, according to the method of Lagrange multipliers, we need to find all solutions to

$$4y = \lambda(2x) \tag{E1}$$
$$4x = \lambda(4y) \tag{E2}$$
$$x^2 + 2y^2 - 1 = 0 \tag{E3}$$

Equation (E1) gives $y = \frac{1}{2}\lambda x$. Substituting this into equation (E2) gives

$$4x = 2\lambda^2 x \qquad \text{or} \qquad 2x\big(2 - \lambda^2\big) = 0.$$

So (E2) is satisfied if either $x = 0$ or $\lambda = \sqrt{2}$ or $\lambda = -\sqrt{2}$.

- If $x = 0$, then (E1) gives $y = 0$ too. But $(0,0)$ violates the constraint equation (E3). Note that, to have a solution, *all* of the equations (E1), (E2) and (E3) must be satisfied.
- If $\lambda = \sqrt{2}$, then
  - (E2) gives $x = \sqrt{2}y$ and then
  - (E3) gives $2y^2 + 2y^2 = 1$ or $y^2 = \frac{1}{4}$ so that
  - $y = \pm\frac{1}{2}$ and $x = \sqrt{2}y = \pm\frac{1}{\sqrt{2}}$.
- If $\lambda = -\sqrt{2}$, then
  - (E2) gives $x = -\sqrt{2}y$ and then

- (E3) gives $2y^2 + 2y^2 = 1$ or $y^2 = \frac{1}{4}$ so that
- $y = \pm\frac{1}{2}$ and $x = -\sqrt{2}y = \mp\frac{1}{\sqrt{2}}$.

We now have four possible values of (x, y), namely $\left(\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$, $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{2}\right)$, $\left(\frac{1}{\sqrt{2}}, -\frac{1}{2}\right)$ and $\left(-\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$. They are the four corners of a single rectangle. We said that we wanted (x, y) to be the upper right corner, i.e. the corner in the first quadrant. It is $\left(\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$.

Example 56.

Find the ends of the major and minor axes of the ellipse $3x^2 - 2xy + 3y^2 = 4$. They are the points on the ellipse that are farthest from and nearest to the origin.

Solution. Let (x, y) be a point on $3x^2 - 2xy + 3y^2 = 4$. This point is at the end of a major axis when it maximizes its distance from the centre, $(0, 0)$ of the ellipse. It is at the end of a minor axis when it minimizes its distance from $(0, 0)$. So we wish to

Maximize and minimize the distance

$$\sqrt{x^2 + y^2}$$

subject to the constraint

$$g(x, y) = 3x^2 - 2xy + 3y^2 - 4 = 0.$$

Now maximizing/minimizing $\sqrt{x^2 + y^2}$ is equivalent to maximizing/minimizing its square $\left(\sqrt{x^2 + y^2}\right)^2 = x^2 + y^2$.

So we are free to choose the objective function

$$f(x, y) = x^2 + y^2$$

which we will do, because it makes the derivatives cleaner. Again, we use Lagrange multipliers to solve this problem, so we start by finding the partial derivatives.

$$\begin{aligned} &f_x(x, y) = 2x \quad f_y(x, y) = 2y \\ &g_x(x, y) = 6x - 2y \quad g_y(x, y) = -2x + 6y \end{aligned}$$

We need to find all solutions to

$$\begin{aligned} &2x = \lambda(6x - 2y) \\ &2y = \lambda(-2x + 6y) \\ &3x^2 - 2xy + 3y^2 - 4 = 0 \end{aligned}$$

Dividing the first two equations by 2, and then collecting together the x's and the y's gives

$$(1 - 3\lambda)x + \lambda y = 0 \tag{E1}$$

$$\lambda x + (1 - 3\lambda)y = 0 \tag{E2}$$

$$3x^2 - 2xy + 3y^2 - 4 = 0 \tag{E3}$$

To start, let's concentrate on the first two equations. Pretend, for a couple of minutes, that we already know the value of λ and are trying to find x and y.

Note that λ cannot be zero because if it is, (E1) forces $x = 0$ and (E2) forces $y = 0$ and $(0, 0)$ is not on the ellipse, i.e. violates (E3). So we may divide by λ and (E1) gives

$$y = -\frac{1 - 3\lambda}{\lambda} x$$

Subbing this into (E2) gives

$$\lambda x - \frac{(1-3\lambda)^2}{\lambda} x = 0$$

Again, x cannot be zero, since then $y = -\frac{1-3\lambda}{\lambda} x$ would give $y = 0$ and $(0,0)$ is still not on the ellipse.

So we may divide $\lambda x - \frac{(1-3\lambda)^2}{\lambda} x = 0$ by x, giving

$$\begin{aligned} \lambda - \frac{(1-3\lambda)^2}{\lambda} = 0 &\Leftrightarrow (1-3\lambda)^2 - \lambda^2 = 0 \\ &\Leftrightarrow 8\lambda^2 - 6\lambda + 1 = (2\lambda - 1)(4\lambda - 1) = 0 \end{aligned}$$

We now know that λ must be either $\frac{1}{2}$ or $\frac{1}{4}$. Subbing these into either (E1) or (E2) gives

$$\begin{aligned} \lambda = \frac{1}{2} \quad &\Longrightarrow \quad -\frac{1}{2}x + \frac{1}{2}y = 0 \quad \Longrightarrow \quad x = y \\ &\Longrightarrow \quad 3x^2 - 2x^2 + 3x^2 = 4 \quad \Longrightarrow \quad x = \pm 1 \\ \lambda = \frac{1}{4} \quad &\Longrightarrow \quad \frac{1}{4}x + \frac{1}{4}y = 0 \quad \Longrightarrow \quad x = -y \\ &\Longrightarrow \quad 3x^2 + 2x^2 + 3x^2 = 4 \quad \Longrightarrow \quad x = \pm\frac{1}{\sqrt{2}} \end{aligned}$$

We now have $(x,y) = \pm(1,1)$, from $\lambda = \frac{1}{2}$, and $(x,y) = \pm\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ from $\lambda = \frac{1}{4}$. The distance from $(0,0)$ to $\pm(1,1)$, namely $\sqrt{2}$, is larger than the distance from $(0,0)$ to $\pm\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, namely 1. So the ends of the minor axes are $\pm\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ and the ends of the major axes are $\pm(1,1)$. Those ends are sketched in the figure on the left below. Once we have the ends, it is an easy matter to sketch the ellipse as in the figure on the right below.

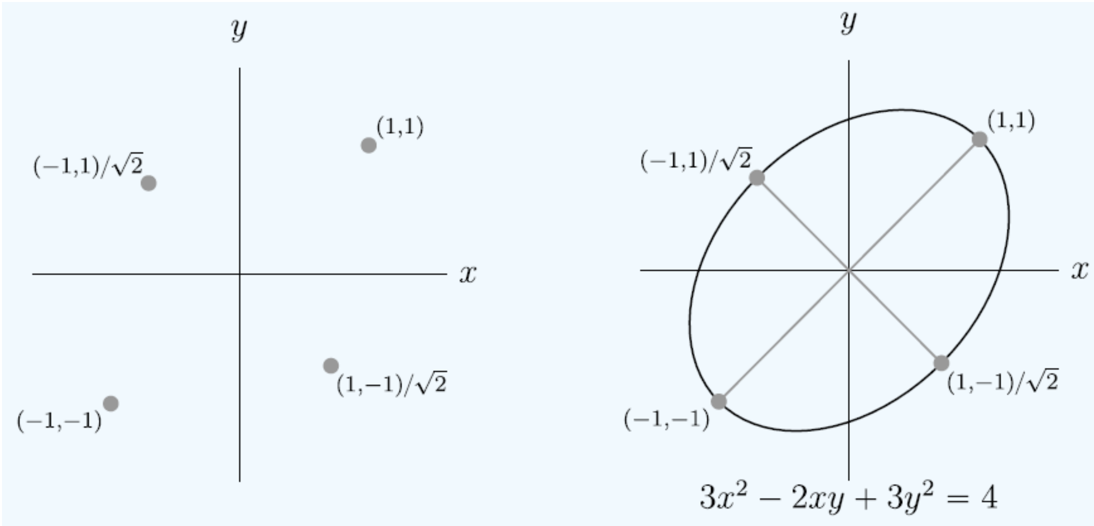


Figure 26

Example 57.

Find the point on the sphere $x^2 + y^2 + z^2 = 1$ is farthest from $(1, 2, 3)$.

Solution. As before, we simplify the algebra by maximizing the square of the distance rather than the distance itself. So we are to maximize

$$f(x, y, z) = (x - 1)^2 + (y - 2)^2 + (z - 3)^2$$

subject to the constraint

$$g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0.$$

We have

$$\begin{aligned} &f_x = 2(x - 1), \quad f_y = 2(y - 2), \quad f_z = 2(z - 3) \\ &g_x = 2x, \qquad\quad g_y = 2y, \qquad\quad g_z = 2z. \end{aligned}$$

We need to find all solutions to the system:

$$\begin{aligned}
&2(x-1)=\lambda(2x) \Leftrightarrow x=\frac{1}{1-\lambda}\\
&2(y-2)=\lambda 2y \Leftrightarrow y=\frac{2}{1-\lambda}\\
&2(z-3)=\lambda 2z \Leftrightarrow z=\frac{3}{1-\lambda}\\
&x^2+y^2+z^2-1=0.
\end{aligned}$$

Solving x, y, z from the first three equations, we get

$$\frac{1+4+9}{(1-\lambda)^2}-1=0 \Rightarrow 1-\lambda=\pm\sqrt{14}.$$

We can then substitute these two values of λ back into the expressions for in terms of x, y, z to get the two points

$$\frac{1}{\sqrt{14}}(1,2,3), \qquad -\frac{1}{\sqrt{14}}(1,2,3).$$

We thus obtain two vectors:

one from $\frac{1}{\sqrt{14}}(1,2,3)$ to $(1,2,3)$ –

$$\left[1-\frac{1}{\sqrt{14}}\right](1,2,3)$$

and the other one from $-\frac{1}{\sqrt{14}}(1,2,3)$ to $(1,2,3)$ –

$$\left[1+\frac{1}{\sqrt{14}}\right](1,2,3).$$

Clearly, the first vector is shorter than the second one. Hence the nearest point is $\frac{1}{\sqrt{14}}(1,2,3)$ and the farthest point is $-\frac{1}{\sqrt{14}}(1,2,3)$.