Unit 3: Partial Derivatives

Dr.P.M.Bajracharya

August 15, 2023

Contents

1 Multivariable Calculus in Data Science

You need to know some basic calculus in order to

- understand how functions change over time (derivatives) and
- 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. For it, it's important to know about *gradients*.

Gradient

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.

GIF1

- As you see this ball follows a path that ends at the bottom of the bowl.
- We also say that the ball is descending in the bottom of the bowl.
- Here, the red lines are gradient of the bowl and the blue line is the path of the ball.

- As the path of the ball's slope is decreasing, it is called as *gradient descent*.

Gradient Descent

Gradient Descent in *machine learning models* and *neural networks* is an optimization algorithm for finding a local minimum of a differentiable function.
It is simply used to find the values of a function's parameters (coefficients) that minimize error function as far as possible.

Suppose that a model involves a function given by

$$y = mx + b,$$

where

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

The goal of gradient descent is to find the parameters m and b that minimizes the difference (error) between the predicted output of the model and improves the model's performance.

Cost Function

The difference (error) between the predicted output of the model and the actual output is known as the **cost function** of the model.

Thus, the gradient descent is commonly-used to train *machine learning models* and *neural networks*.

Training data helps these models learn over time, and the cost function within gradient descent is used to monitor the error in predictions of an ML model, estimating its accuracy with each iteration of parameter updates. Until the function is close to or equal to zero, the model will continue to adjust its parameters to yield the smallest possible error.

Once machine learning models are optimized for accuracy, they can be powerful tools for artificial intelligence (AI) and computer science applications.

Let's 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 John's and Eve's 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.

In theory this is it for gradient descent.

gif2

However, to calculate and model, gradient descent requires calculus and now we can see importance of calculus in machine learning.

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 (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, we'll iterate through each (x, y) point in our data set and sum the square distances between each point's y-value and the candidate line's y-value (computed at $mx + b$). It's 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.

You have seen that to calculate slope, we use *differentiation*.

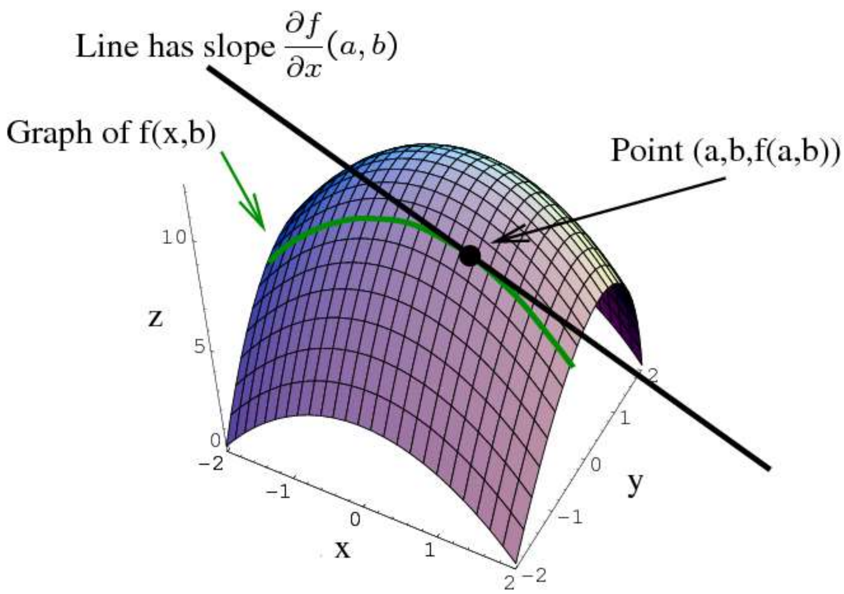


Figure 1: The graph of a function $z = f(x,y)z = f(x,y)$ is a surface, and fixing $y = by = b$ gives a curve (shown in green). The partial derivative $\partial f \partial x(a,b) \partial f \partial x(a,b)$ is the slope of the tangent line to this curve at the point where $x = ax = a$.

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.

Learning Rate

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. It's possible to have a problem with local minima that a gradient search can

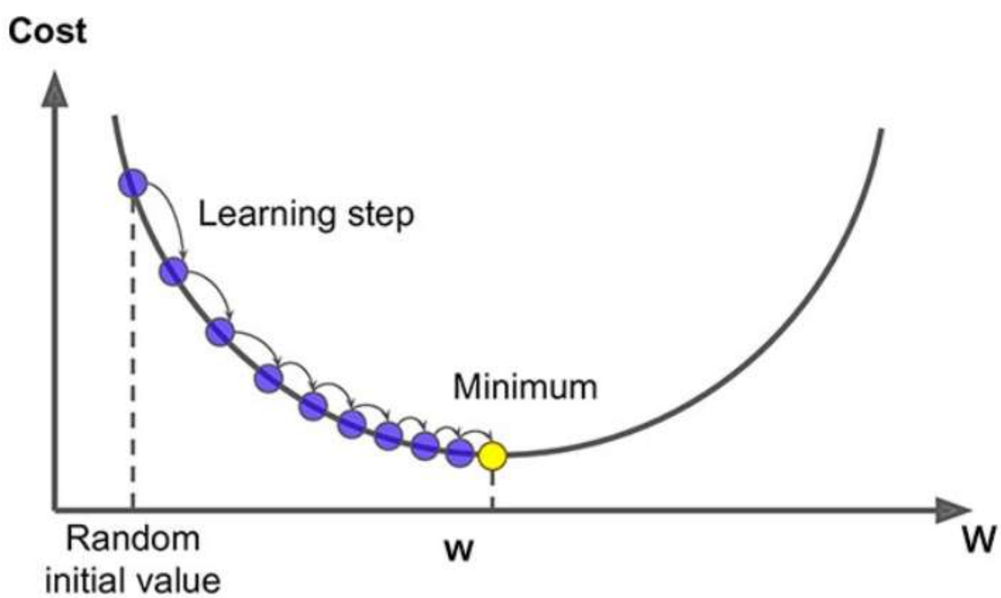


Figure 2: The graph of a function $z = f(x, y)z = f(x, y)$ is a surface, and fixing $y = by = b$ gives a curve (shown in green). The partial derivative $\partial f \partial x(a, b) \partial f \partial x(a, b)$ is the slope of the tangent line to this curve at the point where $x = ax = a$.

get stuck in. There are several approaches to mitigate this (e.g., stochastic gradient search).

gif5

Convergence – We didn't 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

2 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.

2.1 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.

A function of two variables

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.

2.1.1 Determining the domain and the range of a function of two variables

As with functions of one variable, the most common way to describe a function of two variables is with an equation, and unless it is otherwise restricted, you can assume that

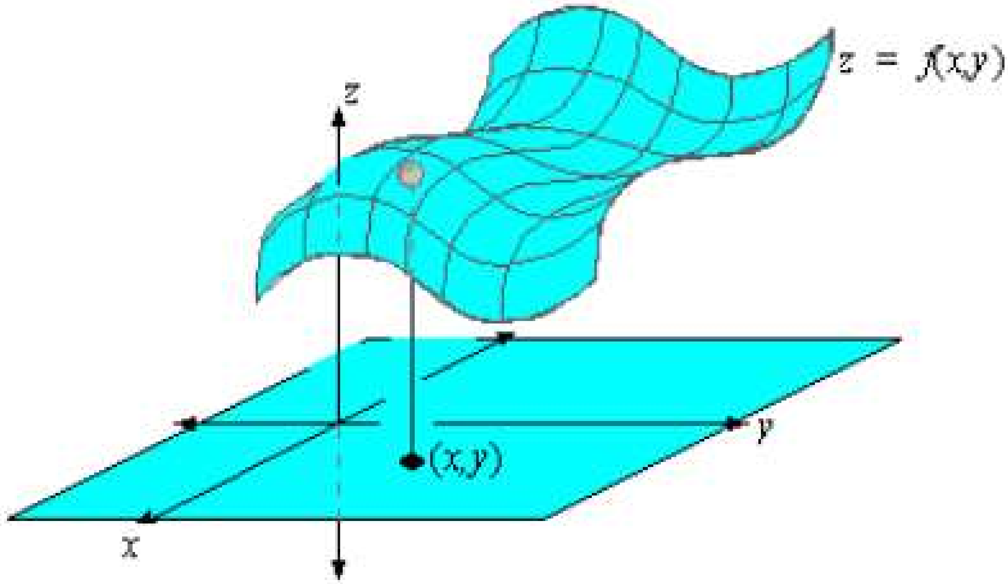


Figure 3: The domain of a function of two variables consists of ordered pairs (x, y).

Domain

The domain is the set of all points for which the equation is defined.

For instance,

1. Consider a function:

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

   What is the domain of the function f?

2. Consider a function:

$$g(x, y) = \ln xy.$$

   What is the domain of the function g?

Ans.: 1. The entire -plane. 2. The set of all points in the first and third quadrants.

Example 1 (Domains). Find the domain of each of the following functions

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

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

(c) $h(x, y) = \dfrac{\sqrt{x^2 + y^2 - 9}}{x}$

Solution.

(a) The function f is defined for all points $(x, y) \in \mathbb{R}^2$. Hence the domain of the function f is $\mathbb{R}^2$.

(b) The function g is defined for all points (x, y) and

$$x^2 + y^2 \leq 9.$$

So, the domain is the set of all points lying on or inside the circle $x^2 + y^2 = 9$. That is,

$$\text{Domain} = \{(x, y) \in \mathbb{R}^2 |\ x^2 + y^2 \leq 9\}.$$

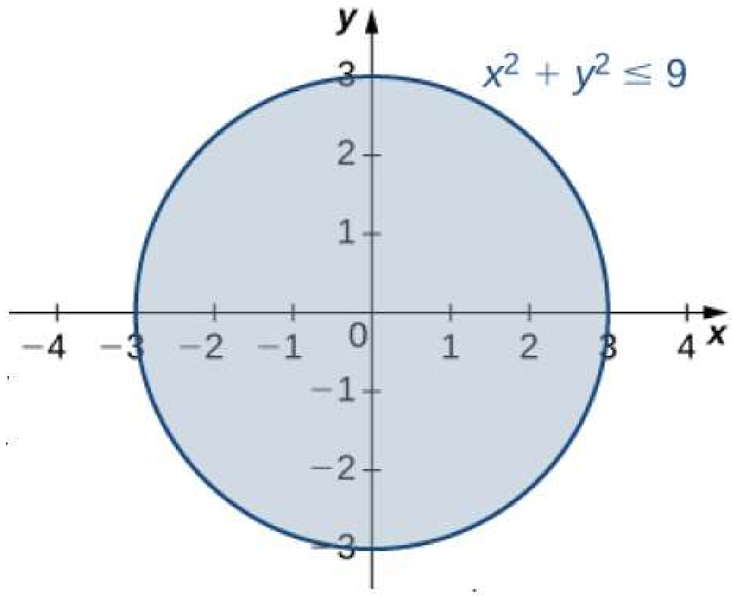


Figure 4: The domain of the function
$g(x, y) = \sqrt{9 - x^2 - y^2}$.

(c) The function h is defined for all points (x, y) such that $x \neq 0$ and

$$x^2 + y^2 \geq 9.$$

So, the domain is the set of all points lying on or outside the circle $x^2+y^2 = 9$ except those points on the y-axis. That is,

$$\text{Domain} = \{(x, y) \in \mathbb{R}^2 | \ x^2 + y^2 \geq 9, x \neq 0\}.$$

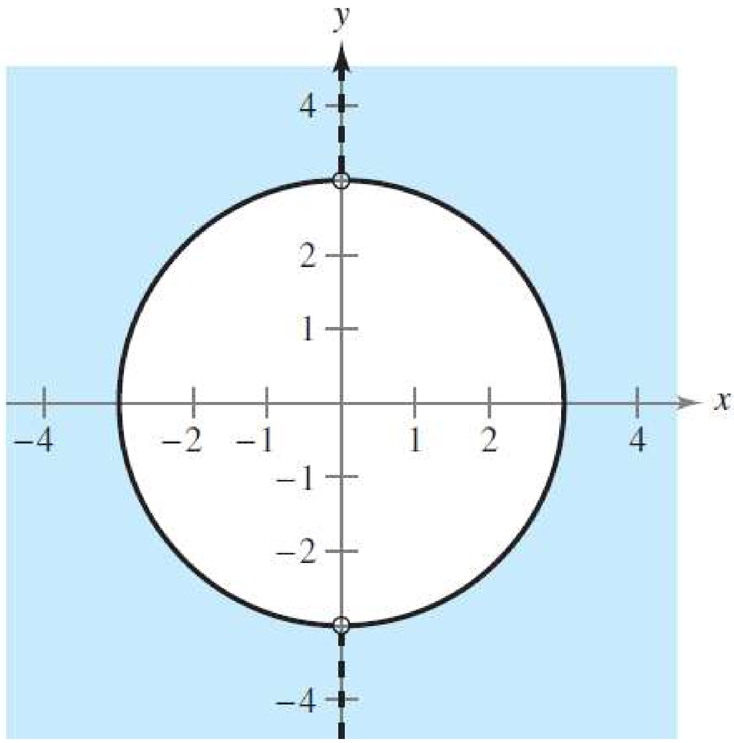


Figure 5: The domain of the function
$h(x, y) = \frac{\sqrt{x^2+y^2-9}}{x}$.

Functions of several variables can be combined in the same ways as functions of single variables. For instance, you can form the sum, difference, product, and quotient of two functions of two variables as follows.

Algebraic operations

$$(f \pm g)(x, y) = f(x, y) \pm g(x, y)$$
$$(fg)(x, y) = f(x, y)g(x, y)$$
$$\frac{f}{g}(x, y) = \frac{f(x, y)}{g(x, y)}$$

We cannot form the composite of two functions of several variables. We can, however, form the composite function $(g \circ f)(x, y)$

where g is a function of a single variable and f is a function of two variables.

The composite function $g \circ f$

$$(g \circ f)(x, y) := g(f(x, y)).$$

The domain of this composite function consists of all (x, y) in the domain of f such that $f(x, y)$ is in the domain of g.

Example 2. A function $h(x, y) = \sqrt{16 - x^2 - 4y^2}$. It can be viewed as the composite of the function f of two variables given by

$$f(x, y) = 16 - x^2 - 4y^2$$

and the function g of a single variable given by

$$g(u) = \sqrt{u}.$$

The domain of this function is the set of all points lying on or inside the ellipse $x^2 + 4y^2 = 16$.

2.1.2 Graph of a Function of Two Variables

As with functions of a single variable, you can learn a lot about the behavior of a function of two variables by sketching its graph.

The graph of a function of two variables is the set of all points (x, y, z) for which $z = f(x, y)$ and (x, y) is in the domain of f. This graph can be interpreted geometrically as a surface in space.

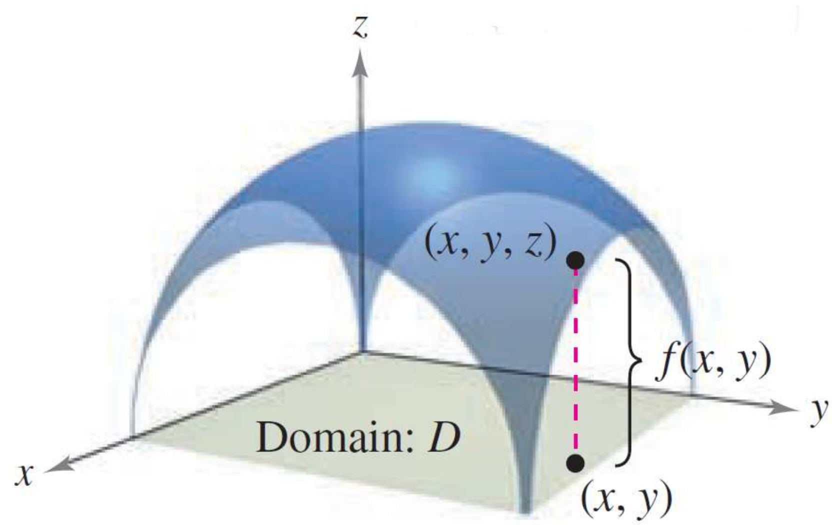


Figure 6: The surface $(x, y, f(x, y))$.

In the above Figure, note that the graph of is a surface whose projection onto the xy-plane is the domain of f. To each point (x, y) in D there corresponds a point (x, y, z) on the surface, and, conversely, to each point (x, y, z) on the surface there corresponds a point (x, y) in D.

Example 3 (Range)**.** Find the range of each of the following functions

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

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

(c) $f(x, y) = \sqrt{16 - 4x^2 - y^2}$.

Describe the graphs of f in each case.

Solution.

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

Range $= \mathbb{R}$.

b The domain D implied by the equation of f is the set of all points (x, y) such that

$$9 - x^2 - y^2 \geq 0.$$

So, D is the set of all points lying on or inside the circle

$$x^2 + y^2 = 9.$$

<u>Range of f</u>

We have

$$z = \sqrt{9 - x^2 - y^2} = \sqrt{9 - (x^2 + y^2)}.$$

Since $x^2 + y^2 \geq 0$, we obtain

$$0 \leq z \leq \sqrt{9} = 3.$$

This is the range of f.

<u>Graph of f</u>

A point (x, y, z) is on the graph of f if and only if

$$\begin{aligned} z &= \sqrt{9 - x^2 - y^2} \\ i.e. \quad z^2 &= 9 - x^2 - y^2 \\ i.e. \quad x^2 + y^2 + z^2 &= 0, \end{aligned}$$

which is the equation of a sphere with centre at the origin and of radius 3. Also, we know that $0 \leq z \leq 4$. Hence the graph of f is the upper half of the sphere, as shown in the figure given below.

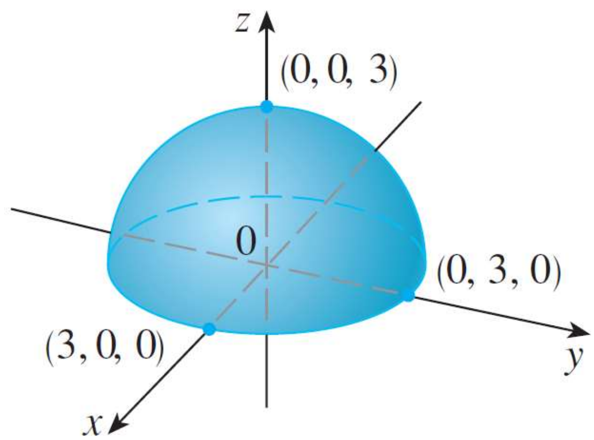


Figure 7: The graph of $f(x, y) = \sqrt{9 - x^2 - y^2}$
is the upper half of a sphere.

c The domain D implied by the equation of f is the set of all points (x, y) such that

$$16 - 4x^2 - y^2 \geq 0.$$

So, D is the set of all points lying on or inside the ellipse

$$\frac{x^2}{4} + \frac{y^2}{16} = 1.$$

Range of f

We have

$$z = \sqrt{16 - 4x^2 - y^2} = \sqrt{16 - (4x^2 + y^2)}.$$

Since $4x^2 + y^2 \geq 0$, we obtain

$$0 \leq z \leq \sqrt{16} = 4.$$

This is the range of f.

Graph of f

A point (x, y, z) is on the graph of f if and only if

$$z = \sqrt{16 - 4x^2 - y^2}$$

$$i.e. \quad z^2 = 16 - 4x^2 - y^2$$

$$i.e. \quad \frac{x^2}{4} + \frac{y^2}{16} + \frac{z^2}{16} = 1,$$

where $0 \le z \le 4$. Hence the graph of f is the upper half of an ellipsoid, as shown in the figure given below.

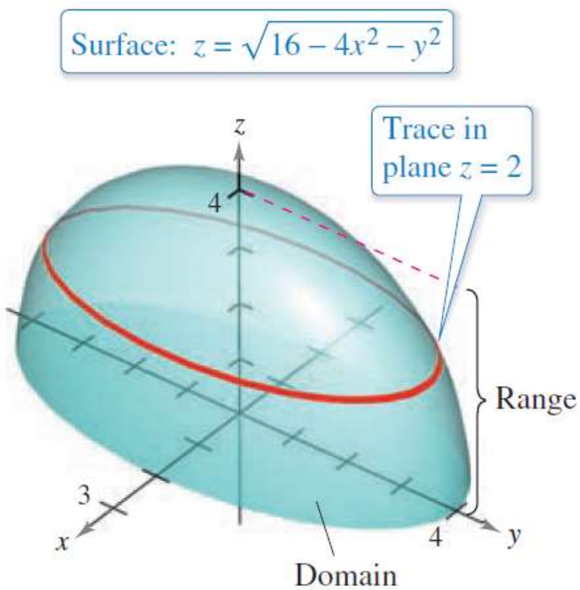


Figure 8: The graph of $f(x, y) = \sqrt{16 - 4x^2 - y^2}$ is the upper half of an ellipsoid.

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

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

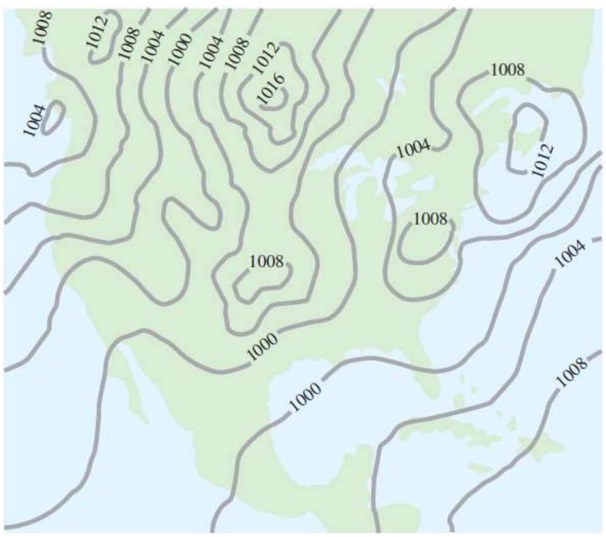
Problem 3. *A profit function for a hardware manufacturer is given by* $f(x, y) = 16 - (x-3)^2 - (y-2)^2$, *where* x *is the number of nuts sold per month (measured in thousands) and* y *represents the number of bolts sold per month (measured in thousands). Profit is measured in thousands of dollars.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

2.1.3 Contour maps

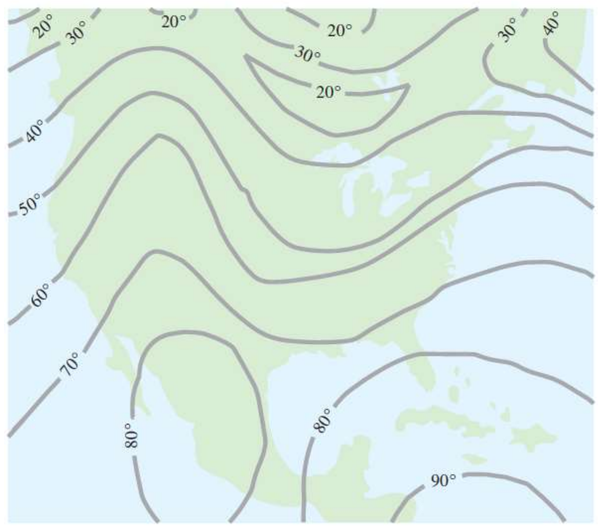
Another method for visualizing functions, borrowed from map-makers, is a contour map on which points of constant elevation are joined to form *contour lines*, or *level curves*.

Contour map

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.

Example 4. Create a contour map for the surface

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

corresponding to $c = 0, 1, 2, ..., 8$.

Solution.

For $c_1 = 0$, we have

$$\sqrt{64 - x^2 - y^2} = 0$$
$$\Longrightarrow x^2 + y^2 = 64.$$

Hence the level curve for f using $c_1 = 0$ is given by

$$x^2 + y^2 = 8^2.$$

For $c_2 = 1$, we have

$$\sqrt{64 - x^2 - y^2} = 1$$
$$\Longrightarrow x^2 + y^2 = 63.$$

Hence the level curve for f using $c_2 = 1$ is given by

$$x^2 + y^2 = (\sqrt{63})^2.$$

For $c_3 = 2$, we have

$$\sqrt{64 - x^2 - y^2} = 2$$
$$\Longrightarrow x^2 + y^2 = 62.$$

Hence the level curve for f using $c_3 = 2$ is given by

$$x^2 + y^2 = (\sqrt{62})^2.$$

Similarly, we find level curves for f using $c_4 = 3, c_5 = 4, c_6 = 5, c_7 = 6, c_8 = 7$.

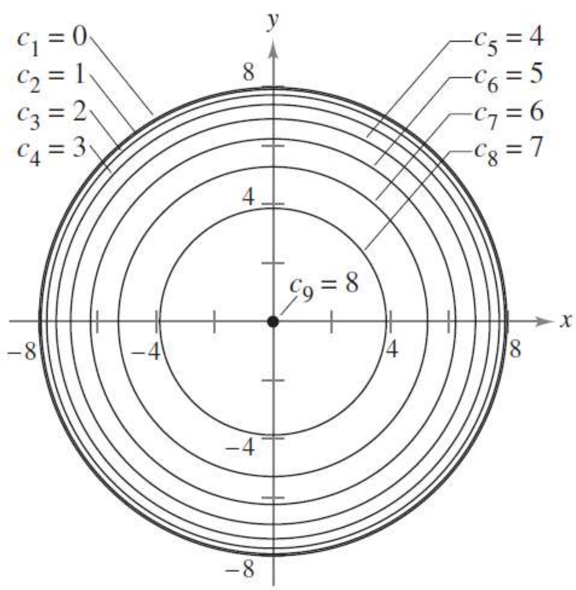
For $c_9 = 8$, we have

$$\begin{aligned} &\sqrt{64 - x^2 - y^2} = 8 \\ \Longrightarrow\ & x^2 + y^2 = 0 \\ \Longrightarrow\ & x = 0,\ y = 0. \end{aligned}$$

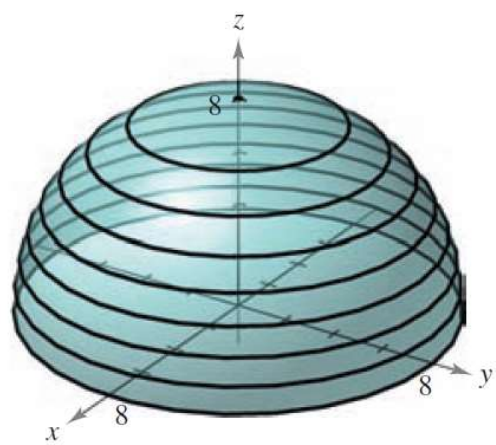
Hence the level curve for f using $c_9 = 8$ is given by

$$x = 0,\ y = 0.$$

It is a degenerate circle.



Contour map of the function $f(x, y) = \sqrt{64 - x^2 - y^2}$, using $c = 0, 1, 2, ..., 8$.



A Hemisphere with level curves.

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. *Create a contour map for the surface*

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

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

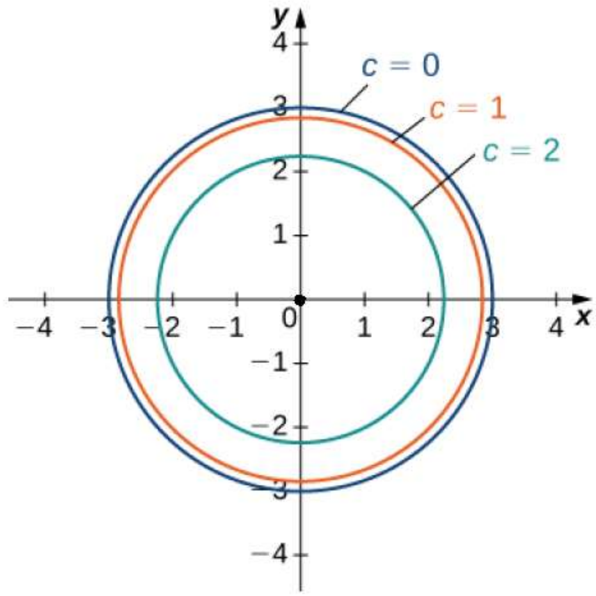


Figure 9: Level curves of the function
$g(x, y) = \sqrt{9 - x^2 - y^2}$, using $c = 0, 1, 2, 3$.

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 *?*

2.2 Functions of Three Variables

Let us take a brief look at functions of three variables.

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

$$f(x, y, z) = \frac{x}{\sqrt{9 - x^2 - y^2 - z^2}}$$

Solution. The function f is defined for all points (x, y, z) such that

$$9 - x^2 - y^2 - z^2 > 0, i.e., x^2 + y^2 + z^2 < 9.$$

Consequently, the domain is the set of all points (x, y, z) lying inside a sphere of radius 3 with center at the origin.

Example 6. 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 7. 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 8. 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 $\sqrt{k}$.

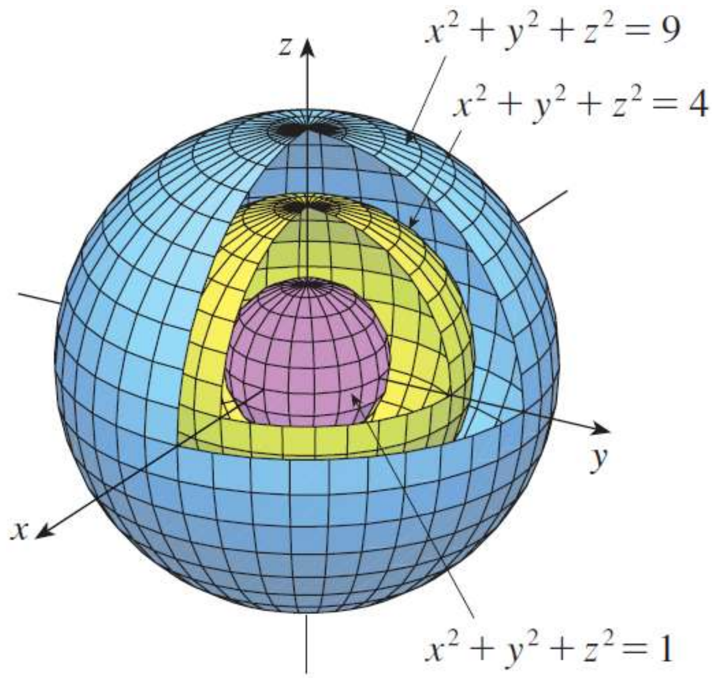


Figure 10: Level surfaces of a sphere.

Problem 1. 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 $k = 2$.

2.3 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 $x_1, x_2, ..., x_n$:

$$C = f(x_1, x_2, ..., x_n) = c_1x_1 + c_2x_2 + \cdot + c_nx_n.$$

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.

3 Limits and continuity

Limits

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

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