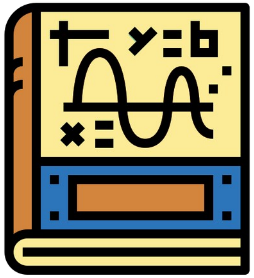
Calculus



Tushar B. Kute,
http://tusharkute.com

Let’s go with an example...

- Sameer and Ajay are traveling in the car ... but the speedometer is broken.
- Ajay: "Hey Sameer! How fast are we going now?"
- Sameer: Wait a minute ...Well in the last minute we went 1.2 km, so we are going... 1.2 km per minute x 60 minutes in an hour = 72 km/h
- Ajay: "No, Sameer! Not our average for the last minute, or even the last second, I want to know our speed RIGHT NOW."
- Sameer: "OK, let us measure it up here ... at this road sign... NOW!"

Let’s go with an example...

- "OK, we were AT the sign for zero seconds, and the distance was ... zero meters!"

- The speed is 0m / 0s = 0/0 = I Don't Know!

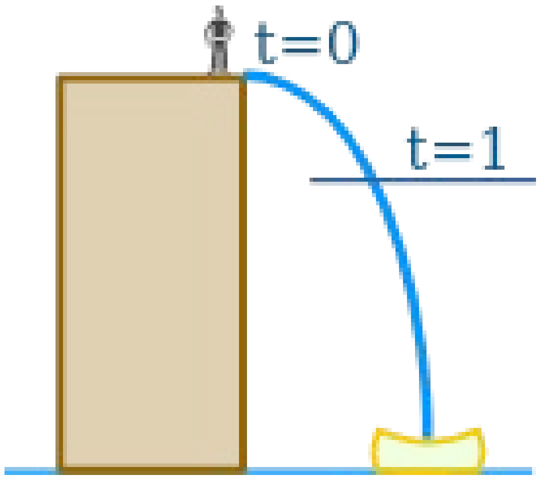
- "I can't calculate it, Ajay! I need to know some distance over some time, and you are saying the time should be zero? Can't be done."

Let’s go with an example...

- That is pretty amazing ... you'd think it is easy to work out the speed of a car at any point in time, but it isn't.
- Even the speedometer of a car just shows us an average of how fast we were going for the last (very short) amount of time.

How About Getting Real Close

- But our story is not finished yet!
- Sameer and Ajay get out of the car, because they have arrived on location. Sameer is about to do a stunt:
- Sameer will do a jump off a 20 m building.
- Ajay, as photographer, asks: "How fast will you be falling after 1 second?"



How About Getting Real Close

- Sam uses this simplified formula to find the distance fallen:

  $d = 5t^2$

  d = distance fallen, in meters

  t = time from jump, in seconds

- (Note: the formula is a simpler version of how fast things fall under gravity: $d = ½gt^2$)
- Example: at 1 second Sameer has fallen

  $d = 5t^2 = 5 \times 12 = 5$ m

How About Getting Real Close

- But how fast is that? Speed is distance over time:

$$\text{Speed} = \frac{\text{distance}}{\text{time}}$$

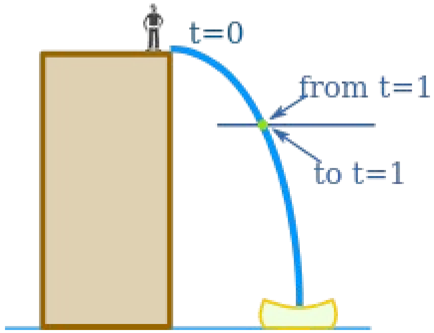
- So at 1 second:

$$\text{Speed} = \frac{5\text{ m}}{1\text{ second}} = 5\text{ m/s}$$

- "BUT", says Ajay, "again that is an average speed, since you started the jump, ... I want to know the speed at exactly 1 second, so I can set up the camera properly."

How About Getting Real Close

- "BUT", says Ajay, "again that is an average speed, since you started the jump, ... I want to know the speed at exactly 1 second, so I can set up the camera properly."



- Well ... at exactly 1 second the speed is:

$$\text{Speed} = \frac{5 - 5 \text{ m}}{1 - 1 \text{ s}} = \frac{0 \text{ m}}{0 \text{ s}} = ???$$

How About Getting Real Close

- So again Sameer has a problem.
- Think about it ... how do we figure out a speed at an exact instant in time?
- What is the distance? What is the time difference?
- They are both zero, giving us nothing to calculate with!

How About Getting Real Close

- But Sam has an idea ... invent a time so short it won't matter.
- Sam won't even give it a value, and will just call it "Δt" (called "delta t").
- So Sam works out the difference in distance between t and $t+\Delta t$

How About Getting Real Close

- At 1 second Sam has fallen

  $5t^2 = 5 \times (1)^2 = 5$ m
- At $(1+\Delta t)$ seconds Sam has fallen

  $5t^2 = 5 \times (1+\Delta t)^2$ m
- We can expand $(1+\Delta t)^2$:

  $(1+\Delta t)^2 = (1+\Delta t)(1+\Delta t)$

  $= 1 + 2\Delta t + (\Delta t)^2$
- So at $(1+\Delta t)$ seconds Sam has fallen

  $d = 5 \times (1+2\Delta t+(\Delta t)^2)$ m

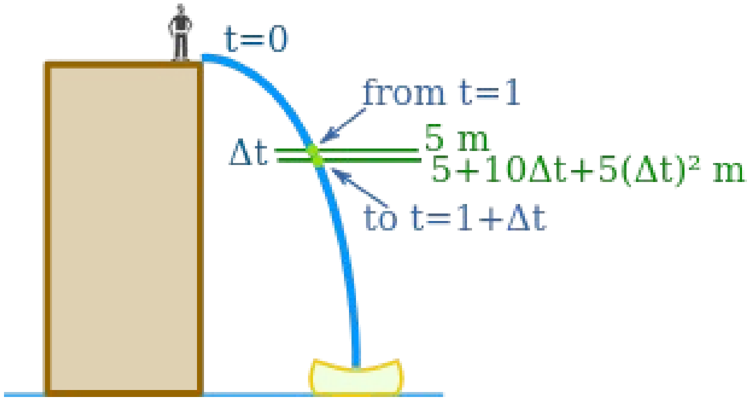
  $d = 5 + 10\Delta t + 5(\Delta t)^2$ m

How About Getting Real Close

- In Summary:

  At 1 second:d = 5 m

  At (1+Δt) seconds:d = 5 + 10Δt + 5(Δt)^2 m



- So between 1 second and (1+Δt) seconds we get:

  Change in d= 5 + 10Δt + 5(Δt)^2 – 5 m

How About Getting Real Close

- Change in distance over time:

$$\text{Speed} = \frac{5 + 10\Delta t + 5(\Delta t)^2 - 5 \text{ m}}{\Delta t \text{ s}}$$

$$= \frac{10\Delta t + 5(\Delta t)^2 \text{ m}}{\Delta t \text{ s}}$$

$$= \mathbf{10 + 5\Delta t} \text{ m/s}$$

- So the speed is 10 + 5Δt m/s, and Sam thinks about that Δt value … he wants Δt to be so small it won't matter … so he imagines it shrinking towards zero and he gets:

Speed = 10 m/s

Finally

- Wow! Sam got an answer!

- Sameer: "I will be falling at exactly 10 m/s"
- Ajay: "I thought you said you couldn't calculate it?"
- Sameer: "That was before I used Calculus!"

What is calculus?

- Calculus is a branch of mathematics that involves the study of rates of change.
- Before calculus was invented, all math was static: It could only help calculate objects that were perfectly still.
- But the universe is constantly moving and changing. No objects—from the stars in space to subatomic particles or cells in the body—are always at rest. Indeed, just about everything in the universe is constantly moving.
- Calculus helped to determine how particles, stars, and matter actually move and change in real time.

What is calculus?

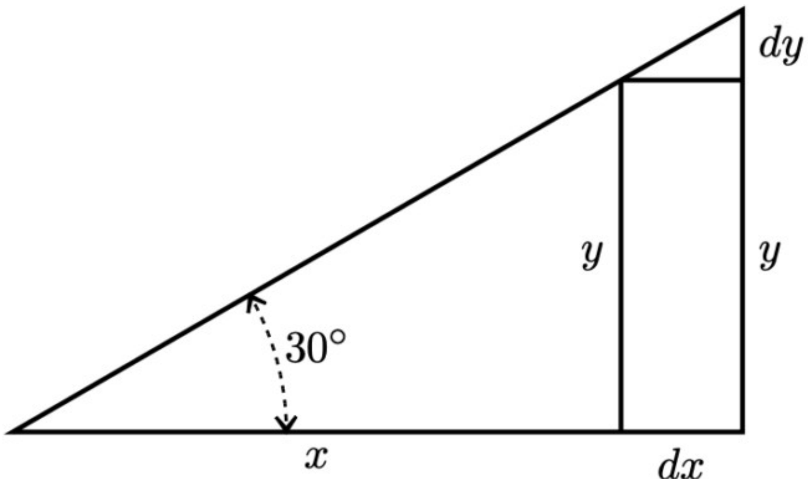
- Calculus is the study of rates of change.
- Gottfried Leibniz and Isaac Newton, 17th-century mathematicians, both invented calculus independently. Newton invented it first, but Leibniz created the notations that mathematicians use today.
- There are two types of calculus: Differential calculus determines the rate of change of a quantity, while integral calculus finds the quantity where the rate of change is known.

What is calculus?

- The word Calculus comes from Latin meaning "small stone".
  - Differential Calculus cuts something into small pieces to find how it changes.
  - Integral Calculus joins (integrates) the small pieces together to find how much there is.
- Differential Calculus and Integral Calculus are like inverses of each other, similar to how multiplication and division are inverses

Derivative

- Let's say that we increase x by a small amount, which we denote dx. And assume that this change in dx increases y by the quantity dy.
- We can visualise this as a right-angled triangle in which x forms the base and y the height.



Derivative

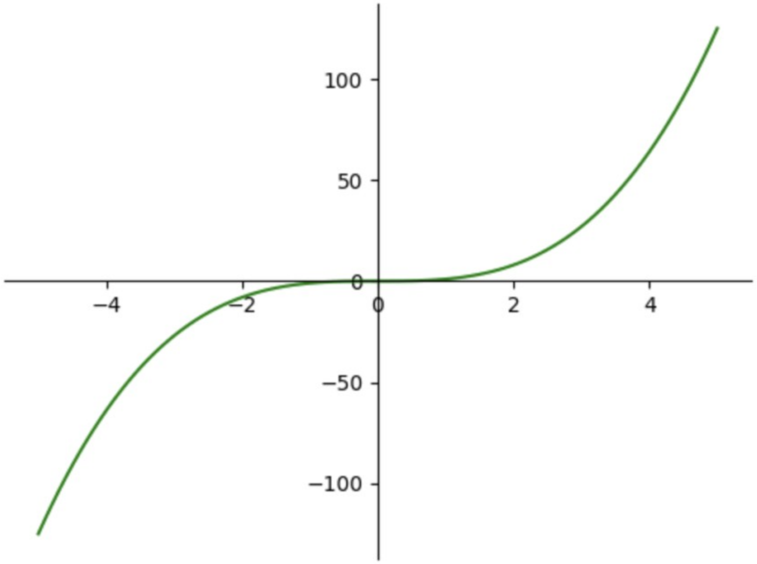
- If we add dx to x, it will cause an increase in y by dy. This corresponding change in dy can be described as a ratio.

$$\frac{dy}{dx}$$

- The expression dy describes a rise in y, whereas dx describes the run (the change in x by dx), which is why this ratio is known as the “rise over run”. In geometric terms, “rise over run” is the slope of our hypotenuse.

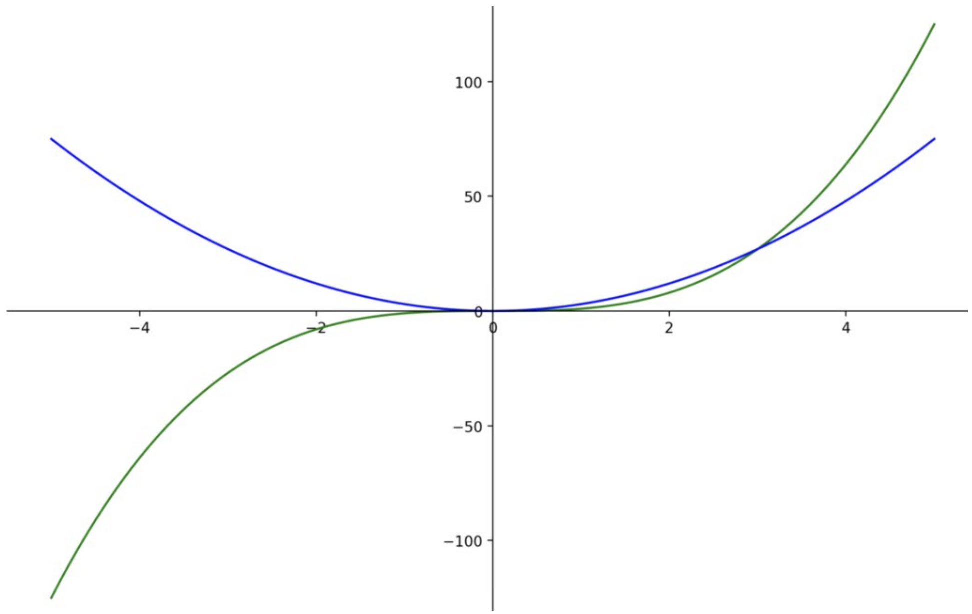
Derivative

- Well, in this case, our hypotenuse is just a straight line. The ratio between dy and dx is the same at every point. It is also equivalent to the ratio of y to x.
- But what do we do if instead of the hypotenuse in the previous image, we need to find the slope of a non-linear graph like this one?



Derivative

- Now the slope and therefore our “rise over run” changes constantly as you move along the line. In other words, the slope itself becomes a non-linear graph (plotted in blue).



Derivative

- Following the origin, the green graph slowly picks up steepness, which also results in a gradual increase in steepness in the blue graph.
- In the real world, the green graph could model the deceleration of a car until it comes to a halt at a traffic light. After the traffic light has turned green, it slowly starts to accelerate again.
- So the green graph captures the actual change in speed while the blue graph records the rate of change in speed at every point in time.

What is Derivative ?

- It should be obvious by now, that we cannot capture the slope of the green graph in a simple slope of y over x.
- To determine the slope of the green graph, we would have to create an infinite number of infinitely small right-angled triangles at every point along the line.
- Now this infinitely small ratio you obtain at every one of these triangles is known as the differential and it is expressed as

$$\frac{dy}{dx}$$

What is Derivative ?

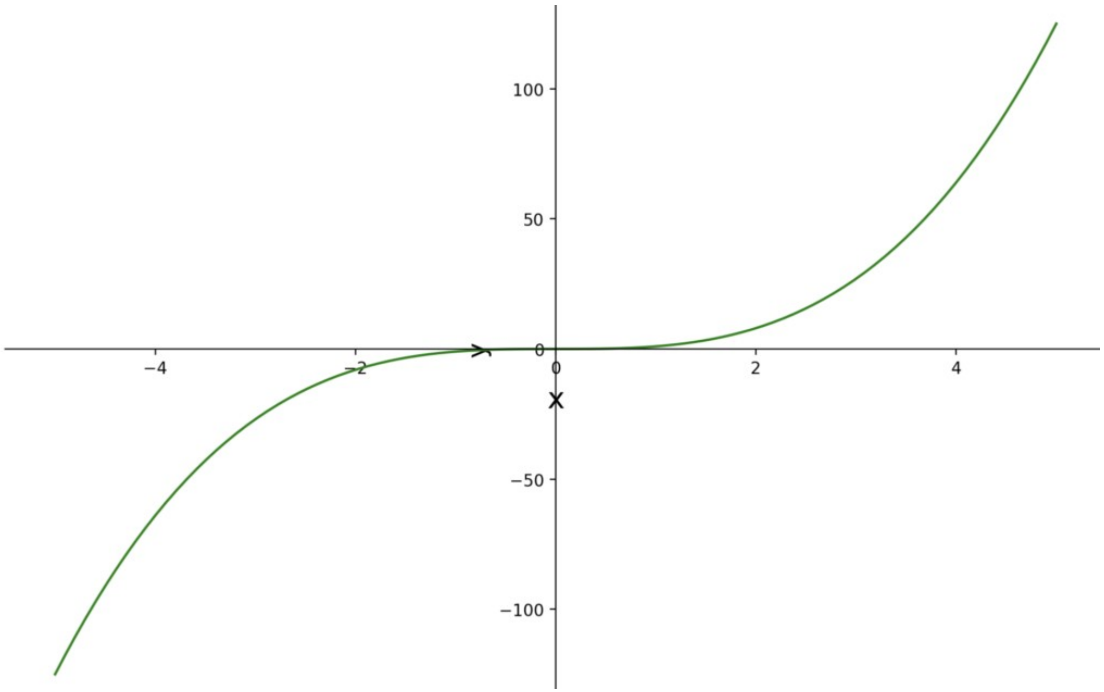
- Here d stands for delta.
- It is known as the differential, and the blue graph that captures the rate of change at every point on the green graph is known as the derivative of the green graph.
- Of course, x and y have to be somehow related to be able to obtain this ratio. We can therefore express the derivative as a function itself.

$$\frac{dy}{dx}$$

How do we got this function?

- Remember in a graph y is usually expressed as a function of x. Our green graph can be expressed as

  $y=x^3$



tusharkute.com

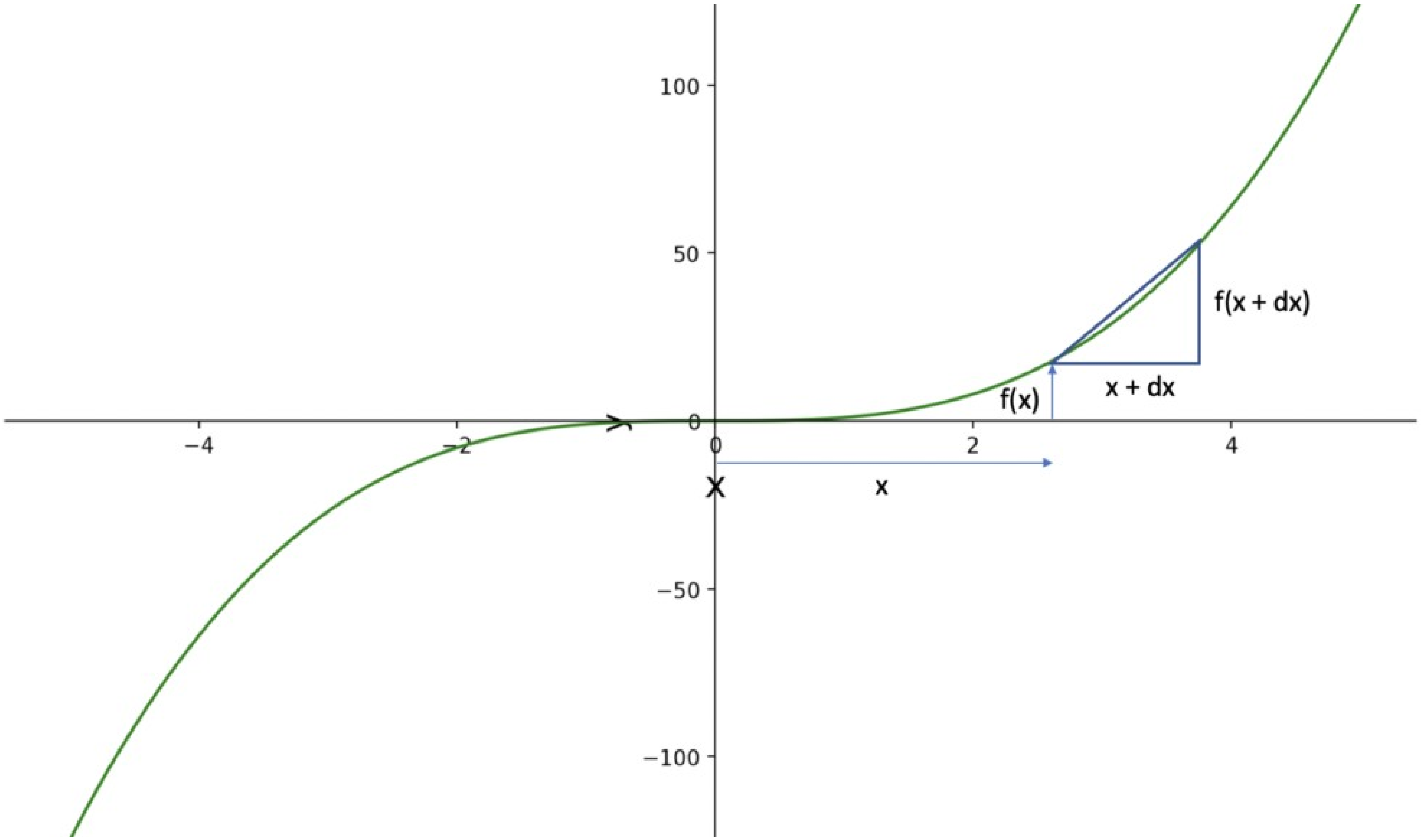
How do we got this function?

- More formally we would say:

  $f(x)=x^3$

- So f(x) is basically a different expression for y in this case.
- You could create a right angled triangle along the graph of f(x), from any point (x, f(x)) to another point (x + dx, f(x + dx)).

How do we got this function?



Rise over run formula

- Accordingly, you would arrive at the slope of this triangle by the following calculation.

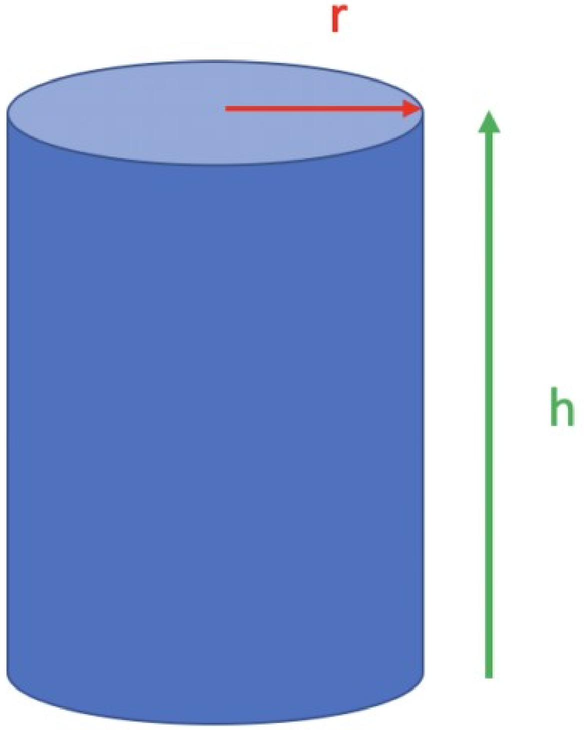
$$\frac{rise}{run} = \frac{dy}{dx} = \frac{f(x+dx) - f(x)}{(x+dx) - x} = \frac{f(x+dx) - f(x)}{dx}$$

- Since the slope is changing continuously in a non-linear graph, we need to make the triangle as small as possible, in fact infinitely small, to arrive at the correct slope at that point.
- As we cannot really calculate the slope of an infinitely small triangle, we can only approximate it by getting x as close to zero as possible.

Partial Derivative

- Many real-life problems in areas such as physics, mechanical engineering, data science, etc., can be described as functions of more than one independent variable.
- How can you know how the change in one particular variable affects the system described by your function?
- This is where partial derivatives come in.
- For instance, the volume of a cylinder can be described as a function of its height h and its radius r.

$V=\pi r^2 h$



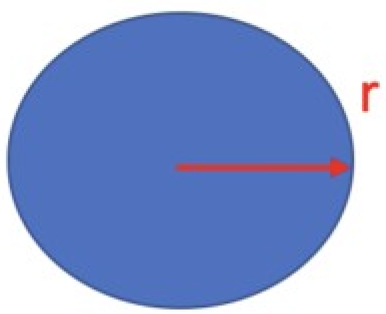
Partial Derivative

- Suppose you need to describe how the volume changes in response to varying just the height while keeping the radius constant.
- To achieve this, you would differentiate the function describing the volume just with respect to h, treating everything else, including r, as a constant. This gives you the following expression.

$$V' = \pi r^2$$

Partial Derivative

- But wait, isn't this the formula for the area of a circle? Indeed, it is.
- That makes perfect sense. If you increase h by an infinitesimally small amount, it is like stacking a circle on top of the cylinder.



Partial Derivative

- The area of the circle is equivalent to the partial derivative of V with respect to h. Formallv we would say.

$$\frac{\partial V}{\partial h} = \pi r^2$$

- Note that ∂ is the partial derivative symbol. You use it instead of d when you are differentiating a multivariate function with respect to one variable.
- If we wanted to find out how V changes if we only increased or decreased r, we would take the partial derivative of V with respect to r.

$$V = \pi r^2 h$$

$$\frac{\partial V}{\partial r} = 2\pi r h$$

What we know?

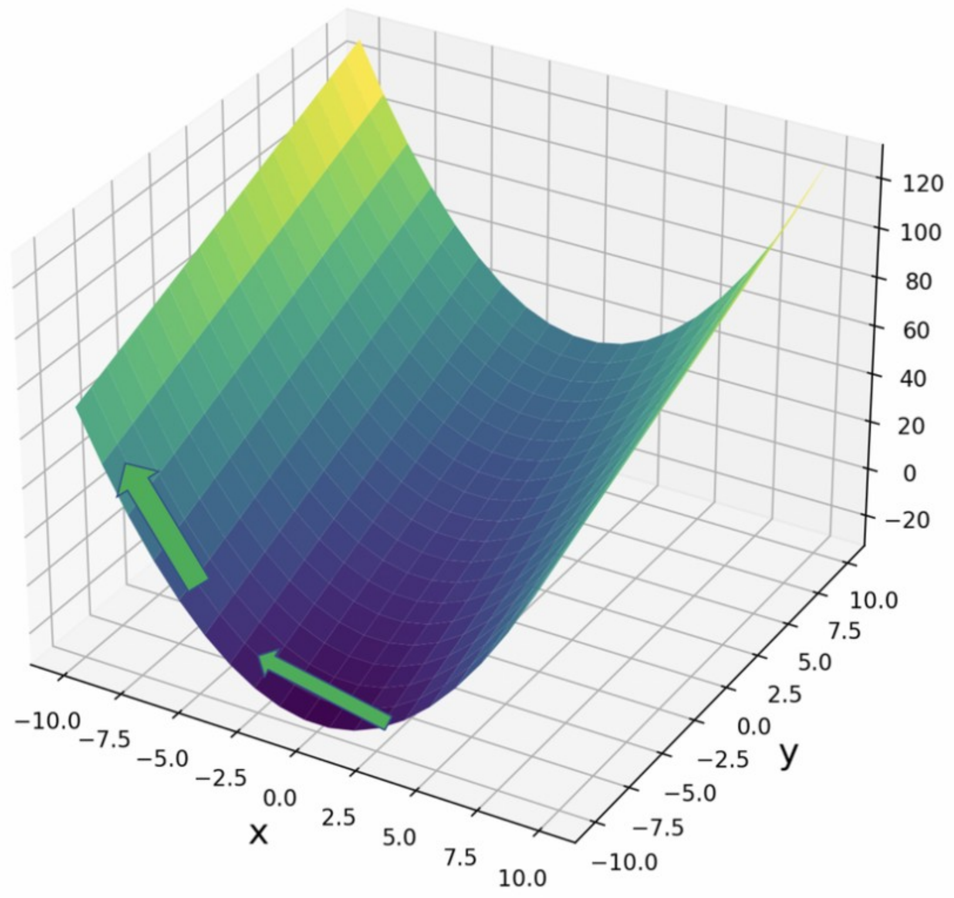
- the derivative is nothing but the slope of a function at a particular point. If we take the multivariate function

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

- The derivative with respect to one variable x will give us the slope along the x dimension.

$$\frac{\partial f(x,y)}{\partial x} = 2x$$

What we know?

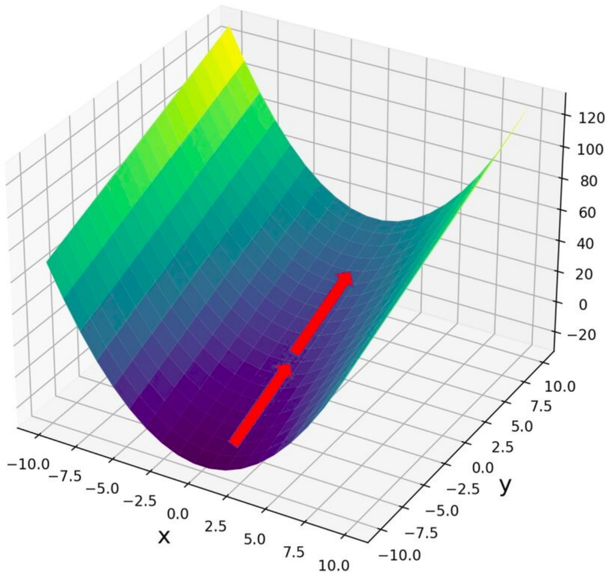

120
100
80
60
40
20
0
−20
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
x
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
y

Vector

- Since x^2 is an exponential term, the slope becomes steeper as we move away from the origin.
- Taking the partial derivative with respect to y gives us the slope along the y dimension.

$$\frac{\partial f(x,y)}{\partial y} = 3$$

3y is a linear term, therefore the slope along the y axis remains constant.



Jacobian Row Vector

- What about the total derivative with respect to x and y? Since we are differentiating with respect to x and y, it is the slope along both dimensions. We can express this as a row vector.

$$df(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} & \frac{\partial f(x,y)}{\partial y} \end{bmatrix}$$

- This is known as the Jacobian matrix. In this simple case with a scalar-valued function, the Jacobian is a vector of partial derivatives with respect to the variables of that function.
- The length of the vector is equivalent to the number of independent variables in the function.

Jacobian Matrix

- The Jacobian of a function of real numbers is a vector. We can expand the definition of the Jacobian to vector-valued functions.

$$f(x) = \begin{bmatrix} f_1(x) \\ \dots \\ f_m(x) \end{bmatrix}$$

- Our function vector has m entries. The resulting Jacobian will be an m×n matrix, where n is the number of partial derivatives.
- Each row m in the matrix contains the partial derivatives corresponding to the equivalent row m in the function vector.

$$\frac{df(x)}{dx} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \dots & & \\ \frac{\partial f_m(x)}{\partial x_1} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

What is use of Jacobian?

- We can use the Jacobian matrix to transform from one vector space to another.
- Furthermore, if the matrix is square, we can obtain the determinant.
- The value of the Jacobian determinant gives us the factor by which the area or volume described by our function changes when we perform the transformation.

Hessian Matrix

- Suppose you are walking around in the hills at night and you would like to find the highest peak.
- You can't see further than a few meters because it is dark. If you followed the direction of the highest slope, you'd eventually end up on a saddle or on a hill, but it might not be the highest point.
- The Hessian gives you a way to determine whether the point you are standing on is, in fact, the highest hill.

Hessian Matrix

- The Hessian matrix is a matrix of the second-order partial derivatives of a function.

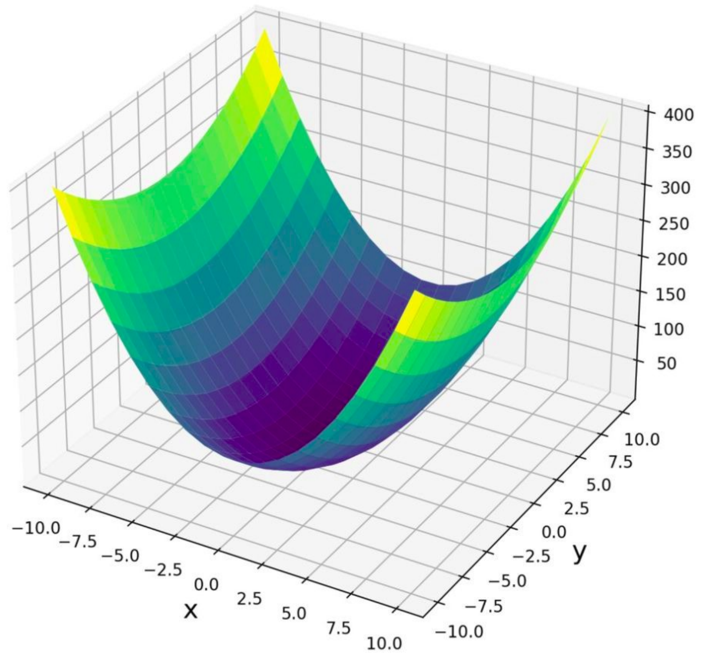
$$df(x,y) = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial xy} \\ \frac{\partial^2 f(x,y)}{\partial xy} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix}$$

- The easiest way to get to a Hessian is to first calculate the Jacobian and take the derivative of each entry of the Jacobian with respect to each variable.
- This implies that if you take a function of n variables, the Jacobian will be a row vector of n entries. The Hessian will be an n×n matrix.

Hessian Matrix

- f you have a vector-valued function with n variables and m vector entries, the Jacobian will be m×n, while the Hessian will be m×n×n .
- Let’s do an example to clarify this starting with the following function.

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



Hessian Matrix

- We first calculate the Jacobian.

  J=[6x 2y]

- Now we calculate the terms of the Hessian.

$$\frac{\partial 6x}{\partial x} = 6 \ and \ \frac{\partial 2y}{\partial x} = 0 \ and \ \frac{\partial 6x}{\partial y} = 0 \ and \ \frac{\partial 2y}{\partial y} = 2$$

$$H = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$$

Hessian Matrix

- Our Hessian is a diagonal matrix of constants. That makes sense since we had to differentiate twice and therefore good rid of all the exponents.
- We can easily calculate the determinant of the Hessian.

  $\det(H) = 6 \times 2 - 0 \times 0 = 12$

Hessian Matrix

- What can we infer from this information?
  - If the first term in the upper left corner of our Hessian matrix is a positive number, we are dealing with a minimum.
  - If the first term in the upper left corner of our Hessian matrix is negative, we are dealing with a maximum.
  - In both cases, the determinant has to be positive
  - If the determinant is negative, the matrix is non-definite. In this case, we might have arrived at a saddle point.

Multivariable Chain Rule

- Remember that the chain rule helps us differentiate nested functions.
- If we have a function f of multiple variables x and y, which are themselves functions of another variable r, we can calculate the total differential.

$$\frac{d}{dr} f(x(r), y(r)) = \frac{\partial f}{\partial x}\frac{dx}{dr} + \frac{\partial f}{\partial y}\frac{dy}{dr}$$

Multivariable Chain Rule

- As we've seen when constructing the Jacobian matrix, then treating x(r) and y(r) as disparate functions, we can write them together in a vector.

$$\vec{v} = \begin{bmatrix} x(r) \\ y(r) \end{bmatrix}$$

- Note that I am writing v with this tiny arrow on top to distinguish it from the other non-vector variables.

Multivariable Chain Rule

- Accordingly, we can also write the derivatives as vectors.

Partial outer derivatives of f with respect to x and y.

$$\frac{\partial f}{\partial(x,y)} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

Nested derivatives of x and y with respect to r.

$$\frac{d(x,y)}{dr} = \begin{bmatrix} \frac{dx}{dr} \\ \frac{dy}{dr} \end{bmatrix}$$

- Now we can write the total derivative of f with respect to the nested variable r as a dot product of the two vectors.

$$\frac{df}{dr} = \frac{\partial f}{\partial(x,y)} \frac{d(x,y)}{dr} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \frac{dx}{dr} \\ \frac{dy}{dr} \end{bmatrix}$$

Example:

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

- Let's first calculate the partial derivatives of f with respect to x, y, and the derivatives for x, y with respect to r.

$$\frac{\partial f}{\partial x} = 4x = 4r^2 - 4$$
$$\frac{\partial f}{\partial y} = 3$$
$$\frac{dx}{dr} = 2r$$
$$\frac{dy}{dr} = 4r$$

Example:

Let’s write them in vector format as a dot product and multiply out.

$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \frac{dx}{dr} \\ \frac{dy}{dr} \end{bmatrix} = \begin{bmatrix} 4r^2 - 4 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2r \\ 4r \end{bmatrix} = 8r^3 + 4r$$

Alternatively, we can eliminate x and y from the start by substituting the appropriate terms of r.

$$\begin{aligned} f(r) &= 2(r^2 - 1)^2 + 3(2r^2 + 3) \\ &= 2r^4 - 4r^2 + 2 + 6r^2 + 6 \end{aligned}$$

Now we can simply differentiate, which gives us the following.

$$\frac{df}{dr} = 8r^3 + 4r$$

Summary

- It resolves to the same term as when we applied the chain rule.
- In this simple case, it is probably faster to use the second method.
- But once you are dealing with many nested variables, the chain rule is a much better and more scalable approach.

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com