Differentiation

Functions

Two types of functions

- Scalar valued single variable function
  - One input, One output $y = f(x)$
- Scalar valued multi variable function
  - Multiple inputs, One output $y = f(x_1, x_2, x_3, \dots)$

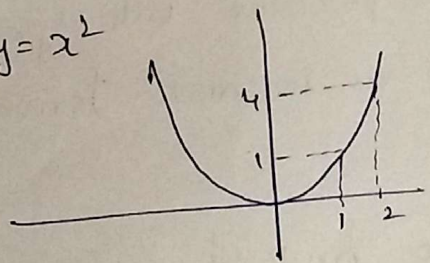
To plot these functions, we have to figure out the number of dimensions needed to plot.

# of dimensions to plot = number of inputs + number of output

So for $y = f(x)$, # of dim = 2,

for $z = f(x,y)$, # of dim = 3

Example $y = x^2$



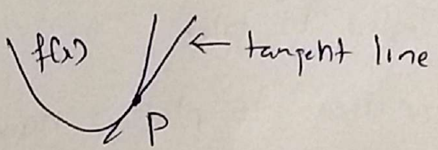
Now here's a deal, I am standing at a pt on the hill. This hill is represented as a fn. I have decided to climb up the hill. To do this, I have to figure out if I have to move forward in x-direction or not. what makes me decide is how the function changes as I move forward. If the function increases as I move forward then, I know I have to continue, else, if the fn decreases as I move forward then I have to change the direction & move in opposite direction.

Differentiation is the one which helps you make this decision.
More technically differentiation will let you know if the fn increases or decreases & by how much as you move forward in the input space.

Differentiation ~~Is also termed as~~ Can also be defined as rate of change of the function.

Slope of a line tells us how steep a line is

Tangent line is a line touching a point on the curve



~~If we combine slope & tangent line it says~~ slope of the tangent line tells how steep is the tangent line drawn at p

In other words slope of the tangent line will speak about the rate of change of the fn at p as we move forward in the input space from p.

Now the question is to find the slope of the tangent line at p. The slope formula is y_2-y_1 / x_2-x_1 so we need two points (x_1,y_1) & (x_2,y_2) on the tangent line, But we just have one point p. Let's introduce Q on the graph which is not on the tangent line. Now if we move Q closer & closer to P, PQ gets parallel to the tangent line at p & thus we get two points on the tangent line PQ^1 where Q^1 is the point very very close to p which is both on the line & on the fn.

It is so close to p that the length of PQ is almost 0.
This is where we say limit $\delta x \to 0$, here δx is PQ' & $\delta x \to 0$ means the length of PQ tends to zero, i.e., Q is very very close to p.
& hence derivative can also be termed as slope of the tangent line at a point of interest & defined as

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x}$$ where δy is the change in y for unit change in x. This unit change is δx

Example

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

$$\frac{\delta y}{\delta x} = \frac{f(x+\delta x) - f(x)}{\delta x} = \frac{(x+\delta x)^2 - x^2}{\delta x}$$

$$= \frac{2x\delta x + (\delta x)^2}{\delta x}$$

$$\frac{\delta y}{\delta x} = 2x + \delta x$$

Applying limit $\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 2x$

$\therefore \frac{dy}{dx} = 2x$ i.e., $\frac{d(x^2)}{dx} = 2x$

Contour Maps

Contour Maps are a way to depict functions with a two-dimensional input and a one-dimensional output.

To plot a graph of such function, we combine the inputs & the output to form a triplet & plot the triplet as a point in the 3-dimensional space.

Collection of all such points form some sort of a surface.

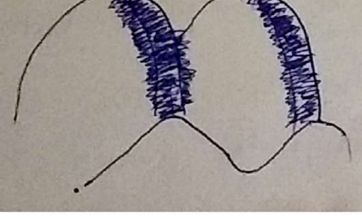
~~Higher~~ Three dimensional surfaces are ~~easier~~ helpful for deeper analysis of a given function. If we want to have a overview, 3-dimensional surfaces are a little too much to grasp. Overview of a function can be best represented using 2-dimensional plots. ~~[illegible] we have to map 3-dim surface on to a 2-dim ~~plot~~ plane.~~

~~Such a representation of 3-dimensional surfaces as a 2-dimensional plot can be achieved using contour maps.~~

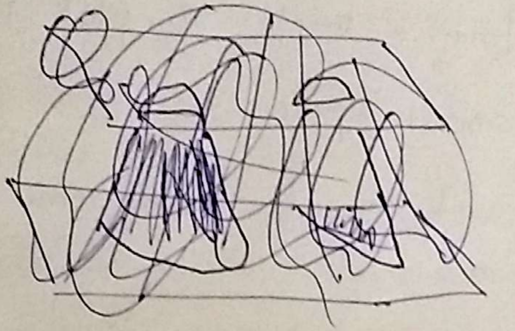
So we have to condense 3-d surface information into a 2-d plot. ~~Once we do that then such 2-d plots gives an overview of the function.~~ This can be achieved using contour-maps. So contour maps are 2-d plots used to get an overview of a function under analysis.

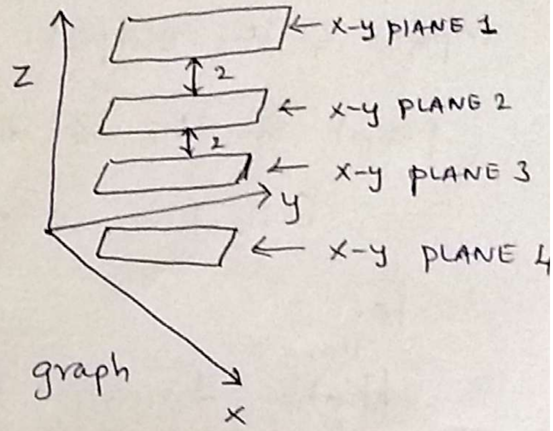
Steps to build contour maps

(i) Start with the graph of the function



(2) Slice the graph with a few evenly-spaced level planes, each of which should be parallel to xy-plane. These planes are equidistant from one another

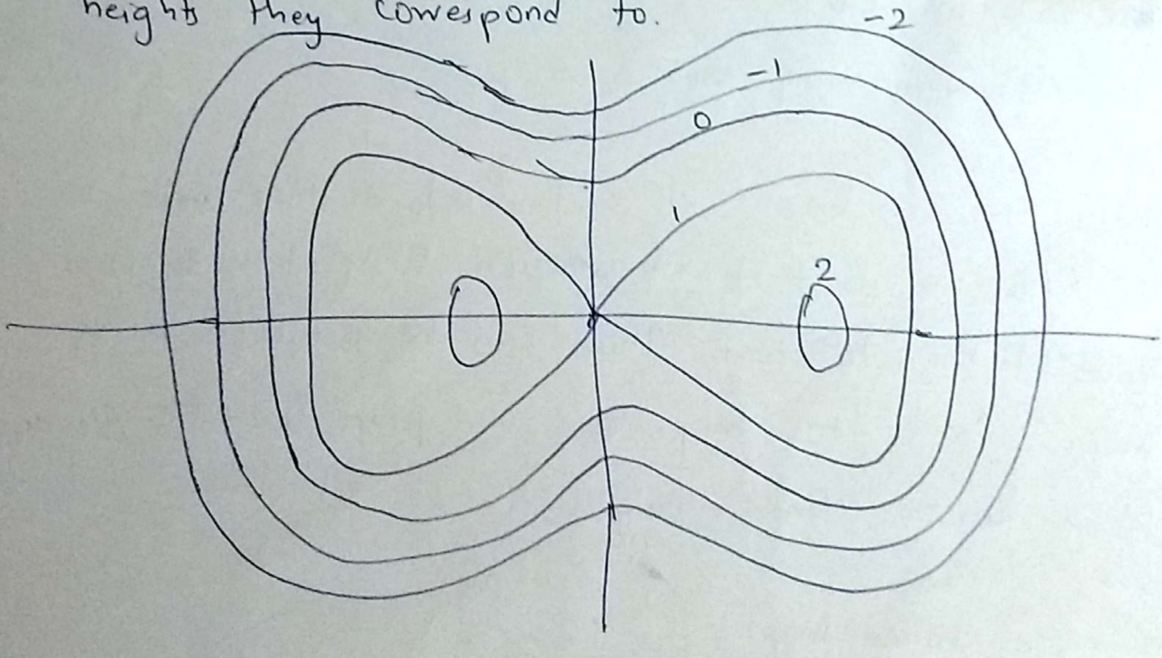




xy plane is slicing the graph

(3) Mark the graph where the planes cut into it

(4) Project these lines onto the xy-plane & label the heights they correspond to.



Here numbers represents the plane number cutting that region. ~~High~~ Bigger the number, higher the plane with bigger z-value. As an example ~~2/6 represents~~ In the above contour maps we have 5 planes cutting through the graph (−2, −1, 0, 1, 2) with −2 representing the plane at lesser height with ~~les~~ smaller z-value & 2 representing the top most plane ~~at~~ with higher z-value.

The lines on a contour map have various names:

(i) Contour lines.

(ii) Level sets

(iii) Iso lines

Depending on what the contour map represents, the iso-prefix might come attached to a number of things.

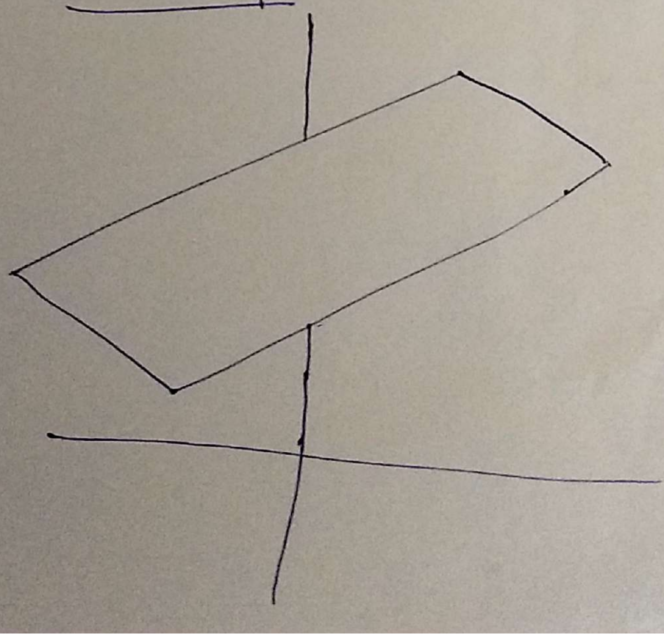
- Isotherm is a line on a contour map for a function representing temperature

- An Isobar is a line on a contour map representing pressure.

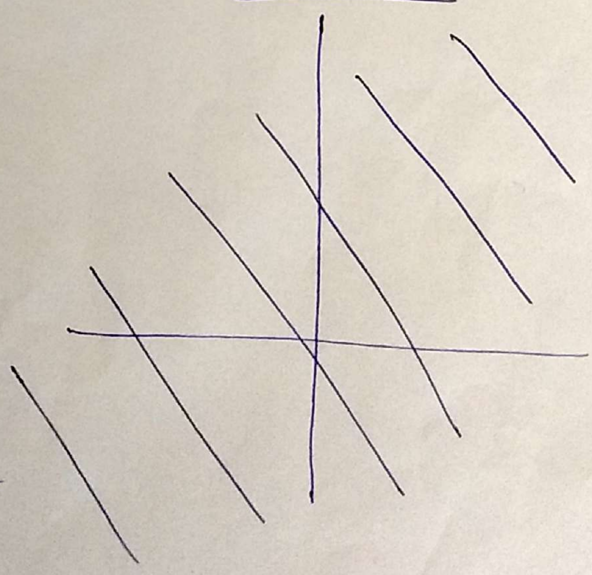
Gaining intuition from a Contour Map

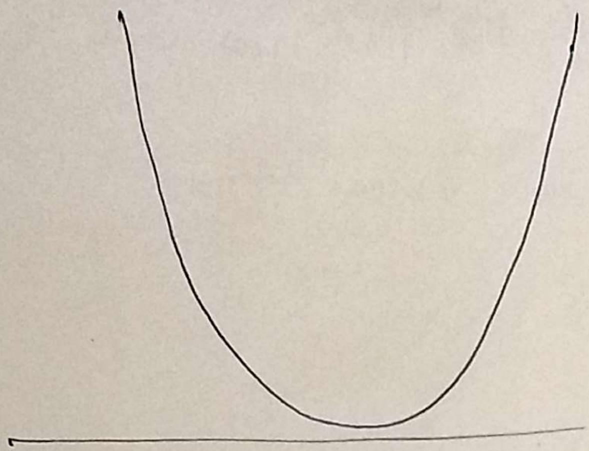
You can tell how steep a portion of your graph is by how close the contour lines are to one another. When they are far apart, it takes a lot of lateral distance to increase altitude.

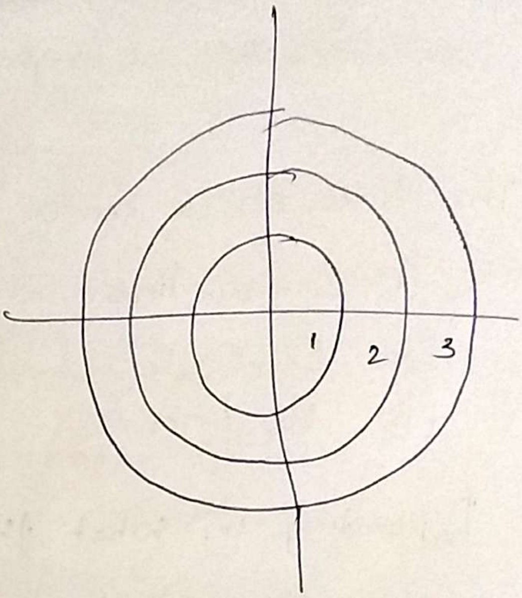
Other Examples



Contour Map

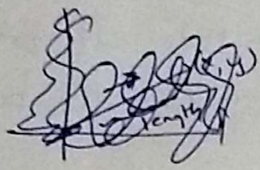


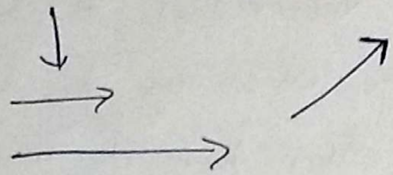



1
2
3

VECTOR FIELDS

① How can we define a moving object on a two dimensional plot. We do it using vectors. The length of the vector (magnitude) indicates speed & the direction indicates which way the object is moving

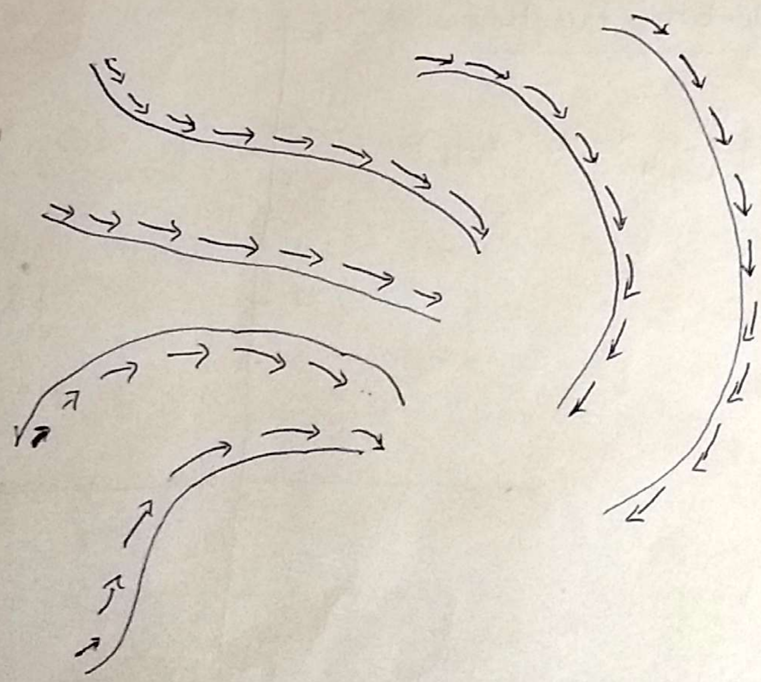




Examples of vector

The description of the vector only tells us its magnitude & direction but not where to draw the vector. ~~The~~ It's upon us to attach the tail of the vector any where we choose to.

Now lets kick it up a notch. what if we want to depict the motion of several objects (like motion of fluid). We have to have several vectors & each such vector corresponding to each object or particle in a fluid.



Looking at the ~~picture~~ diagram we can figure out how the particles in the fluid is flowing.

Such a diagram is called a vector field.

(i) In vector field, the length of each vector in comparison to another says which one has higher magnitude. If one vector is bigger than the other then its magnitude (speed) is bigger than the other.

(ii) The other way of measuring the magnitude in comparison betn vectors is to have different colors depicting different magnitudes.

Example dark blue vectors may represent vectors with higher magnitude than light blue vectors.

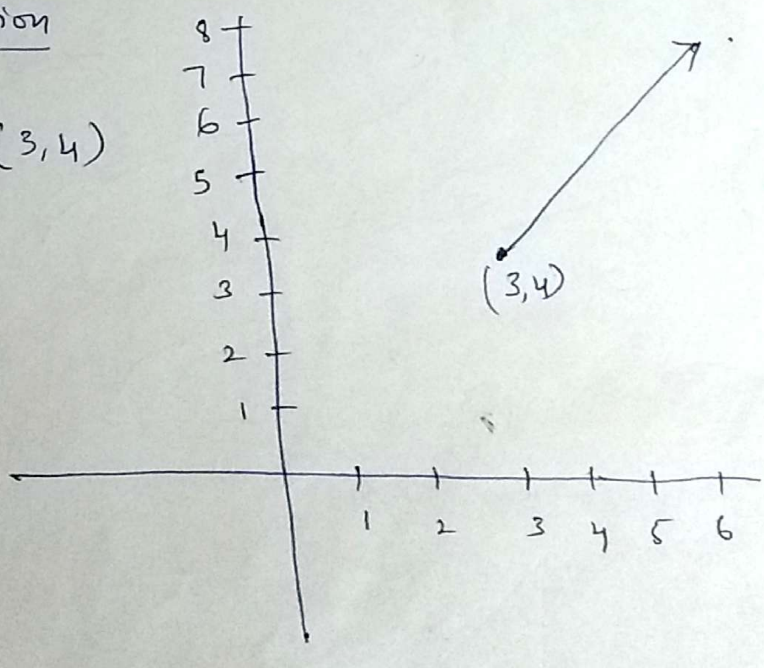
Mathematically a vector field represents multivariable function whose input and output spaces each have the same dimension.

Such a function is also called as vector valued function

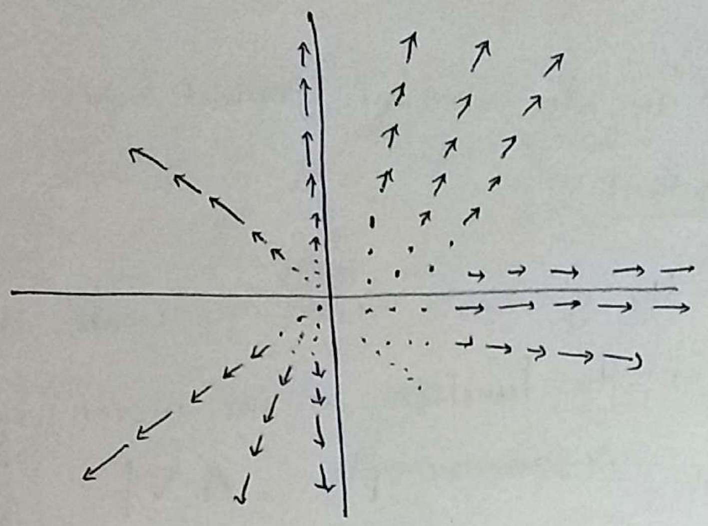
Example

Identity Function

$f(3,4) = (3,4)$



If ~~we~~ you do this for all the points in xy-plane, this is how it looks



In the similar fashion we can depict vector fields in 3D space.

GRADIENT

Set up

Scalar Valued multivariable function $f(x, y, \ldots)$

Vector Valued Function is a multivariable fn with vector as the output

Now Gradient is the vector valued function of a Scalar valued multivariable function.

Interesting isn't it.

So the gradient of f is represented as ∇f & is a vector with partial derivatives of each variable as its component

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \vdots \end{bmatrix}$$

What does these components of the gradient vector indicate.

They indicate how much of a change happens to the function when each component increases by Δ unit.

Vector fields of a fn are also called as gradient fields.

Interpreting the gradient

Each component of the gradient vector represents the amount of change in the function as one moves unit distance on the component axis. ~~As an example~~ $\nabla f = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix}$

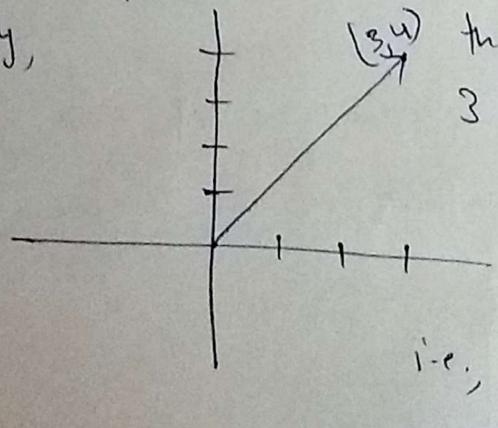
$\frac{\delta f}{\delta x}$ is the partial derivative of f w.r.t x & so measures the change in f ~~as x mo~~ for unit change in x, keeping y constant. IIIly $\frac{\delta f}{\delta y}$ is the change in f for unit changes in y, keeping x constant

Example : If suppose, the gradient of a fn is given as

$$\nabla f = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{i.e, } \frac{\delta f}{\delta x} = 3 \ \& \ \frac{\delta f}{\delta y} = 4$$

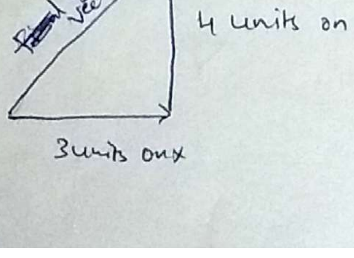
This means ~~at x~~ 'as we move unit length on x axis, keeping y constant, the ~~fn~~ rate of change of fn is 3. IIIly for y-component the rate of change of fn is 4, i.e., it increases by 4.

Pictorially,



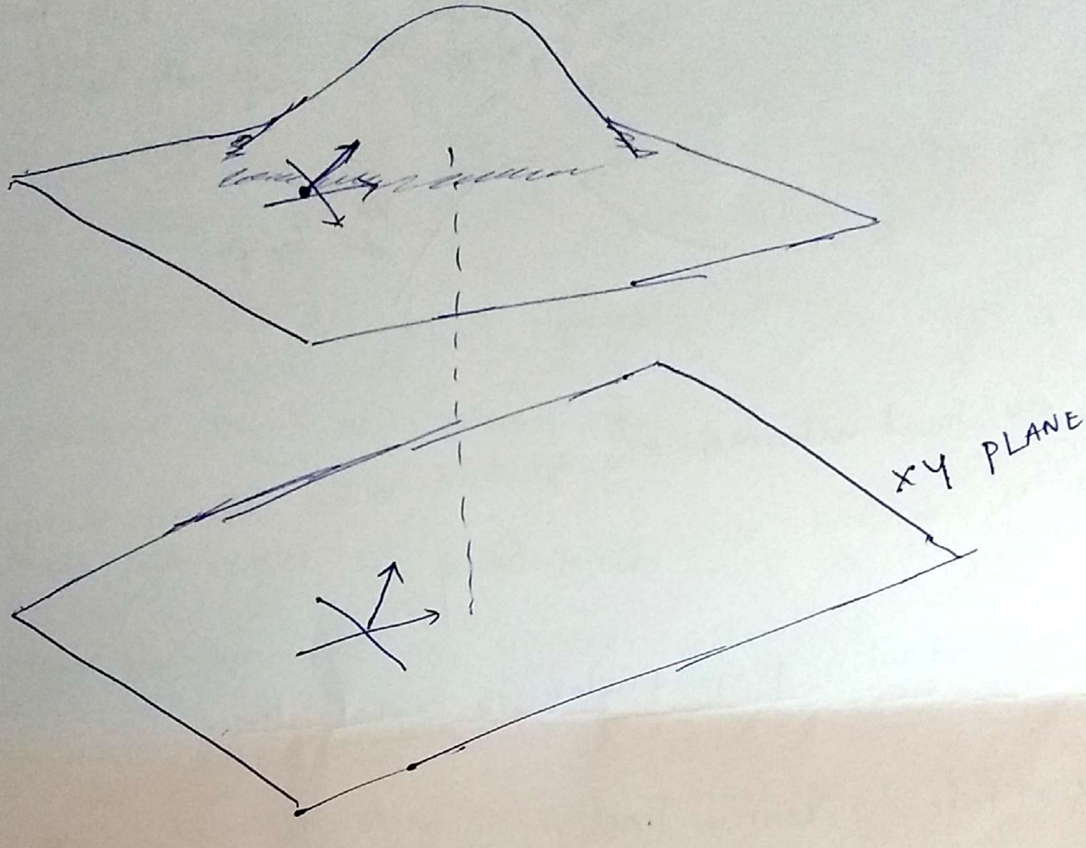
this is the gradient moving 3 units on x & 4 units on y

i.e.,



Point to ponder

Gradient points in the direction of steepest ascent

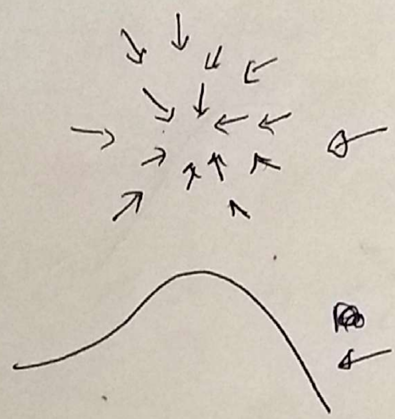


Here we are talking about scalar valued multivariable function. This scalar value can be a +ve value, -ve value or 0.

As we move unit length on the component axes, the ∇f should therefore be either +ve, -ve or 0. If +ve it means it is climbing the ~~steep~~ hill, -ve means getting down the hill & 0 means on the flat land.

So the idea here is that as you keep ~~moving~~ following the gradient vector forward on the ~~component axes finally you will reach the steepest~~ the ascent. locally you will reach the steepest ascent.

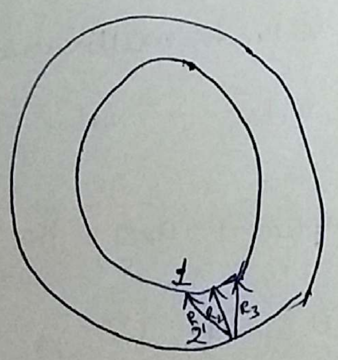
So at the local maxima, the gradients of all the points around the maxima will be pointing towards the peak



 ||ly at the local minima, the gradient vectors at all the points around the minima will point away.

The gradient is perpendicular to contour lines

Let's take 2 contour lines



To go from 2 to 1 we can take many routes (R_1, R_2, R_3) but the shortest amongst R_1, R_2 & R_3 is R_2 which is perpendicular to the contour line. i.e., the line perpendicular to the two parallel lines is the shortest line joining them.

The tools - Partial derivatives, gradient, contour maps etc can be used to optimize and approximate multivariable functions.

Plane

Is a function $f(x,y) = m_1 x + m_2 y + b$

Here $m_1 = \frac{\delta g}{\delta x}$ & $m_2 = \frac{\delta g}{\delta y}$ Here g is another plane Intersecting f

This tells that

1. The partial derivatives of g wrt x & y are constants & they represent m_1 & m_2 ~~which~~

The partial derivative of a plane ~~at~~ wrt x keeping y constant can be imagined as another plane intersecting f at some constant y. This line of intersection has a constant slope & this line is the partial derivative & hence the partial derivative is constant.

Simply put the tangent plane can be got as follows for the following set up

Lets assume we have a function $f(x,y)$ & we want to compute the tangent plane at a point x_0, y_0

So $$\boxed{L_f(x,y) = \frac{\delta f}{\delta x}(x - x_0) + \frac{\delta f}{\delta y}(y - y_0) + f(x_0, y_0)}$$

Local Linearization $= \begin{bmatrix} \cancel{\delta f/\delta x} \\ \cancel{\delta f/\delta y} \\ \cancel{f(x,y)} \end{bmatrix} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} f(x,y) \\ \delta f/\delta x \\ \delta f/\delta y \end{bmatrix}$

~~Local Linearization~~

Properties of Local Linearization

① $L(x,y) = f(x,y)$

② $\frac{\delta L}{\delta x} = \frac{\delta f}{\delta x}$ & ③ $\frac{\delta L}{\delta y} = \frac{\delta f}{\delta y}$

Quadratic approximation

$$Q_f(x,y) = \frac{\delta f}{\delta x}(x-x_0) + \frac{\delta f}{\delta y}(y-y_0) + f(x_0,y_0)$$

$$+ \frac{1}{2}\frac{\delta^2 f}{\delta x^2}(x-x_0)^2 + \frac{1}{2}\frac{\delta^2 f}{\delta y^2}(y-y_0)^2 + \frac{\delta^2 f}{\delta x \delta y}(x-x_0)(y-y_0)$$

Quadratic approximation of a given function will hugs the function quite closely

Hessian Matrix $\cancel{\begin{bmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta x \delta y} \end{bmatrix}} = \begin{bmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta y \delta x} & \frac{\delta^2 f}{\delta y^2} \end{bmatrix}$

Quadratic approximation in vector form

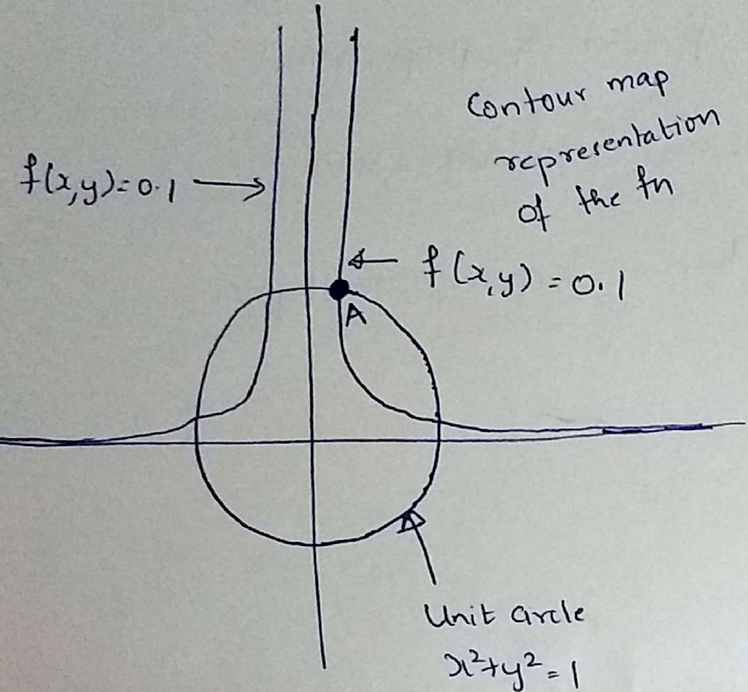
$$Q_f(X) = f(X_0) + \nabla f(X_0)(X-X_0) + \frac{1}{2}(X-X_0)^T H_f(X_0)(X-X_0)$$

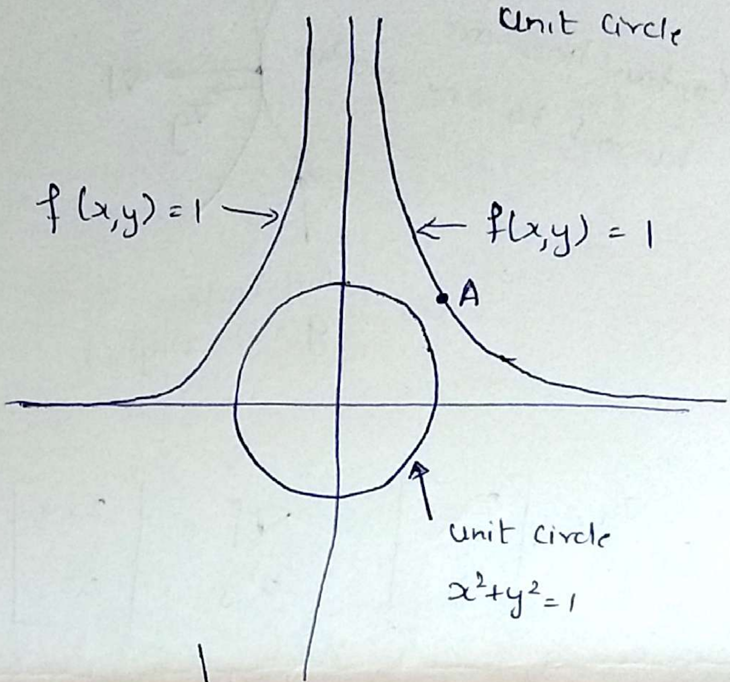
H_f = Hessian Matrix , ∇f is the gradient fn

Lagrange multipliers & constrained Optimization

Maximize a multivariable fn, subject to some constraint.

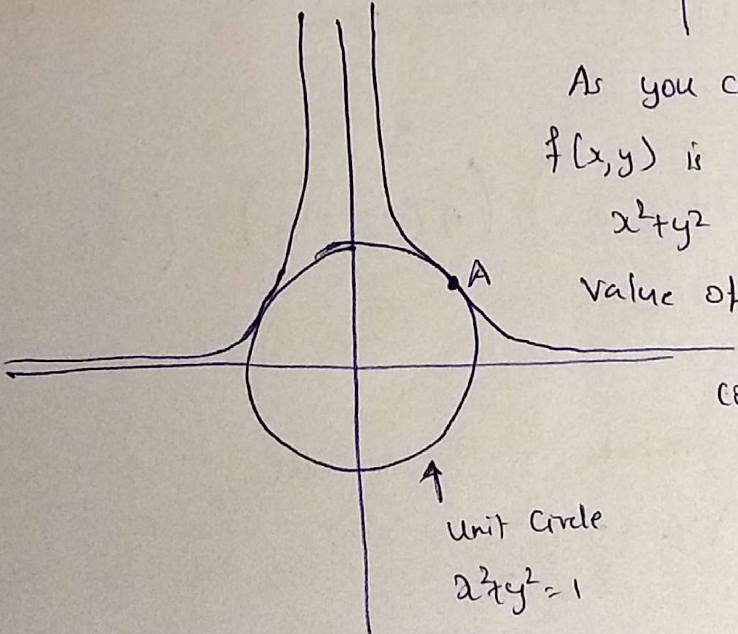
Example : maximize $fn(x,y) = x^2y$ on the set $x^2+y^2=1$ unit circle





At point A, the $f(x,y)$ is intersecting $x^2+y^2=1$ i.e., the values of x & y at that particular point has $f(x,y)=0.1$, also they satisfy $x^2+y^2=1$

Point A doesn't satisfy the constraint $x^2+y^2=1$

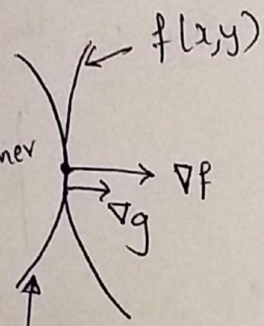


As you can notice, at point A $f(x,y)$ is tangent to unit circle x^2+y^2 & is the maximum value of $f(x,y)$ subject to the constraint.

$f(x,y) = c$

Set c to a value such that $f(x,y)$ kisses the constraint unit circle as shown below

Contour lines are tangent to one another



Unit circle
$g(x,y) = x^2+y^2 = 1$

The gradient vectors ∇f & ∇g are parallel & proportional

$\nabla f(x,y) = \lambda \nabla g(x,y)$

λ here is lagrange multiplier.

$\nabla g = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$ $\quad (x^2+y^2)$

$\nabla f = \begin{bmatrix} 2xy \\ x^2 \end{bmatrix}$ $\quad x^2y$

$\begin{bmatrix} 2xy \\ x^2 \end{bmatrix} = \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix} \Rightarrow$

$2xy = \lambda 2x \Rightarrow y = \lambda$

$x^2 = \lambda 2y \Rightarrow x^2 = 2y^2$

$x^2+y^2 = 1 \Rightarrow 2y^2+y^2 = 1, \; 3y^2 = 1$

4 points where $f(x,y)$ is maximum given the constraint is

$y = \pm\sqrt{1/3} = \lambda$

$x = \pm\sqrt{2/3}$

$(\sqrt{2/3}, \sqrt{1/3})$ $(-\sqrt{2/3}, \sqrt{\tfrac{1}{3}})$ $(\sqrt{2/3}, -\sqrt{1/3})$ $(-\sqrt{2/3}, -\sqrt{1/3})$

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

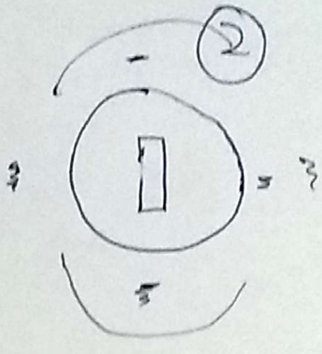
Substitute the values & figure out

$f(\sqrt{2/3}, \sqrt{1/3}) = \frac{2}{3}\sqrt{1/3}$

x & y values maximizing $f(x,y)$ & $x^2+y^2=1$ is

Scanned by CamScanner

Lagrangian



constrained optimization problem

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

Constraint $g(x,y) = x^2 + y^2 = 4$

~~Lagrangia~~ We say at a point (x,y) when the contour lines of both g & f are tangent to one another, we have $\nabla f = \lambda \nabla g$

Lagrangian packages all these into an equation

$$L(x,y,\lambda) = f(x,y) - \lambda\,(g(x,y) - \underbrace{b}_{\text{constant}})$$

Now set $\nabla L = 0$

$$\nabla L = \begin{bmatrix} \frac{\delta L}{\delta x} \\ \frac{\delta L}{\delta y} \\ \frac{\delta L}{\delta \lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{\delta L}{\delta x} = \frac{\delta f}{\delta x} - \lambda \frac{\delta g}{\delta x} = 0 \quad \text{Eq (1)}$$

$$\frac{\delta L}{\delta y} = \frac{\delta f}{\delta y} - \lambda \frac{\delta g}{\delta y} = 0 \quad \text{Eq (2)}$$

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

$$\frac{\delta L}{\delta \lambda} = 0 - \underbrace{(g(x,y) - b) = 0}_{\text{Constraint equation.}} \leftarrow \text{Eq (3)}$$

Significance of λ

$$Max^* = f(x^*, y^*)$$

$$Max^*(b) = f(x^*(b), y^*(b))$$

$$\lambda^* = \frac{dM^*}{db}$$

λ is actually the rate of change of $f(x,y)$ for unit change of constant in the constraint equation $x^2+y^2=b$

Points to ponder

(i) Directional Derivative

$$\nabla_{\vec{v}} f(\vec{a}) = \lim_{h \to 0} \frac{f(\vec{a}+h\vec{v}) - f(\vec{a})}{h}$$

$$= \frac{\nabla f \cdot \vec{v}}{\|\vec{v}\|}$$