UNIT - 4 OPTIMIZATION IN STATISTICS

→ Uses - Resource allocation problem, Neural network training, Travelling salesman problem, Predicting Future Trends

Optimising CPU usage limit, storage capacity to minimize respone time
Constraint : budget, resource

Minimize the CF (MSE for regression problem, Cross entropy for classification)

Finding shortest possible route by each city only once, minimized distance, fuel cost, time

Predict sales & optimize business operation

Optimization

→ Process of finding best solution to a problem, given a set of constraints

→ Key Components include

i) Objective Function - Function to be minimized / maximized $f(x)$

ii) Decision Variable - Often denoted as vector $(x_1, x_2, \dots, x_n)$

iii) Constraints - Inequality $g_i(x) \leq 0$, Equality $h_j(x) = 0$

iv) Feasible Region - Set of all x that satisfy constraints (X)

v) Optimisation methods / algorithms - Algorithm to find the best solution

Try This ☺

4

20 Bananas → 1₹

1 Apple → 1₹

1 Mango → 5₹

100 items with 100₹

each item atleast 1

→ General form :

Optimize $f(x)$ O.F

Subject to $g_i(x) \leq 0$, $i = 1, \dots, m$ inequality constraints

equality constraints $h_j(x) = 0$, $j = 1, \dots p$

decision variable $x \in X$ Feasible region

→ Types of Optimisation

1) Linear Programming - Technique to optimize a linear objective function subject to some constraints
(Resource Planning, Workforce Scheduling, Transportation & Logistics)

2) Integer Programming - Same as linear programming but all / some decision variables must take integer values
(Scheduling Problems, Capital Budgeting)

3) Non-linear Programming - Optimization dealing with problems in which objective f^n and / or constraints are non-linear $(z = 3x_1^2 + 2x_2^2)$
(Neural network, Support Vector Machines, Optimal control problems)

4) Convex Optimization - Minimizing a convex set where any local minimum = global minimum
(Signal Processing, Machine learning)

5) Combinatorial Optimization - Selects best sol^n from finite set of possibilities often involving discrete variables
(Travelling salesman problem, Vehicle routing)

6) Genetic Optimization - Meta-heuristic approach mimicing natural selection to find optimal solution
(Machine Learning Hyperparameter tuning, Engineering Design Problems)

7) Dynamic Programming - Solves complex problems by breaking down into simpler sub-problems once
(Shortest path problems, Inventory Management)

→ Common Optimisation Techniques

i) Gradient Descent - Regression & Neural Networks where goal is to minimize error by iteratively adjusting model parameters

ii) Simplex Method - Used in Linear Programming, where it helps

→ Challenges in Optimization

i) Scalability
ii) Non-convexity
iii) Uncertainity
iv) Multi-Objective Optimization
v) Dynamic Environments
vi) Interpretability

→ Modelling in Optimization

i) Deterministic - All input data & parameters are known with certainity, so, no R.V /uncertainities
ii) Stochastic

i) Mathematical
ii) Statistical
iii) Simulation

examples of Modelling →

	Decision Variable	Objective	Constraints
i) Supply Chain Optimisation	Quantity of Goods	Minimising production cost Maximising prod^n supply	Budget, Time, No. of Vehicles
ii) Portfolio Optimisation in Finance	Allocation of Investment funds across various assets	Maximising profits Minimising risks	Budget, Risk tolerance
iii) Scheduling	Assignment of tasks to workers / machine	Maximising Production Minimising Labourers / Time	Precedence of tasks, Resources, Equipment

Linear Programming Problem

→ The general form of a LPP contains an objective function which is subject to a set of constraints & decision variable in order to satisfy an objective

→ Procedure :

i) Convert all inequality into equations

ii) Plot each equation on the graph

iii) Shade the feasible region

iv) Compute the coordinates of corner points , which represents Feasible Solution

v) Substitute the coordinates into objective function

Q.

Resources			
Product	Labor (Hr/unit)	Clay (Lb/unit)	Profit ($/unit)
Bowl	1	4	40
Mug	2	3	50

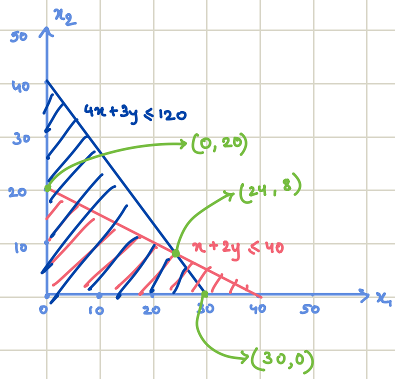
Resource Available : 40 Hrs of Labour per day

120 Lbs of Clay

A. Let x_1 & x_2

x_1 → No. of bowls to be produced per day

x_2 → No. of mugs to be produced per day



Objective Function : $40x_1 + 50x_2$

Resource Constraints : $x_1 + 2x_2 \leqslant 40$

$4x_1 + 3x_2 \leqslant 120$

Non-negativity Constraints : $x_1 \geqslant 0$, $x_2 \geqslant 0$

From graph, Corner points : A(0,20) , B(24,8) , C(30,0) , D(0,0)

A ⇒ $0 + (50 \times 20) = 1000$, B ⇒ $(40 \times 24) + (50 \times 8) = 1360$, C ⇒ $(40 \times 30) + 0 = 1200$, D ⇒ $0 + 0 = 0$

Optimal Solution ⇒ 24 Bowls & 8 Mugs for max profit of 1360 $

Q.

Brand	Resources		
	Nitrogen (Lb/bag)	Phosphate (Lb/bag)	Cost ($/bag)
Super-gro	2	4	6
Crop-quick	4	3	3

Resource Available : Field requires atleast 16 pounds of nitrogen & 24 pounds of phosphate

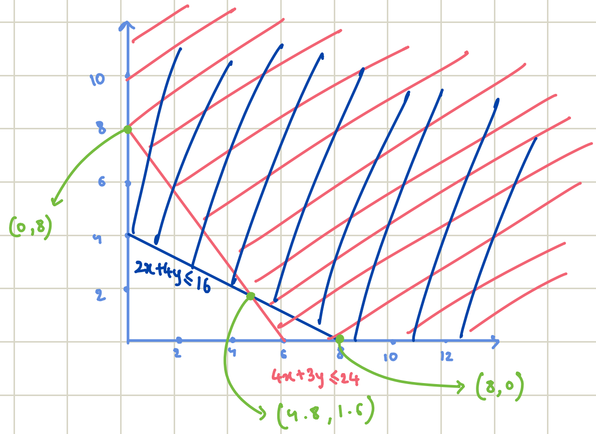
A. Let $x_1 \Rightarrow$ Super-gro

$x_2 \Rightarrow$ Crop-quick

Objective $\Rightarrow$ $6x_1 + 3x_2$

Constraints $\Rightarrow$ $2x_1 + 4x_2 \geq 16$

$4x_1 + 3x_2 \geq 24$



Corner Points $\Rightarrow$ A(0,8) , B(4.8 , 1.6) , C(8,0)

Optimal Sol^n $\Rightarrow$ A $\Rightarrow$ $0 + 3 \times 8 = 24$

B $\Rightarrow$ $(6 \times 4.8) + (3 \times 1.6) = 33.6$

C $\Rightarrow$ $6 \times 8 + 0 = 48$

Least = A

Q. A company manufactures two products, Product A and Product B. The objective is to maximize profit, where Product A gives a profit of $50 per unit, and Product B gives a profit of $30 per unit. However, there are constraints on resources like labor hours and raw materials.

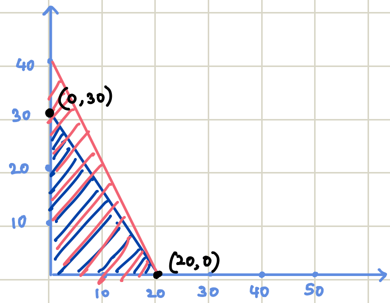
Labor: 2x1 + 1x2 ≤ 40
Materials: 3x1 + 2x2 ≤ 60
Non-negativity: x1, x2 ≥ 0

A. Objective Function : $50x_1 + 30x_2$

Constraints : $2x_1 + x_2 \leq 40$

$3x_1 + 2x_2 \leq 60$

$x_1 , x_2 \geq 0$



Corner Points $\Rightarrow$ $c_1(0, 30)$, $c_2(20, 0)$, $c_3(0, 0)$

Substituting in O.F $\Rightarrow$ $C_1 = 900$, $C_2 = 1000$, $C_3 = 0$

So max profit at $x_1 = 0$ & $x_2 = 20$

Simplex Method

→ $z = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$

Subject to constraints $a_{11} x_1 + a_{12} x_2 + \dots\dots + a_{1n} x_n \quad (\le, =, \ge b_1)$

$\vdots$

$a_{m1} x_1 + a_{m2} x_2 + \dots\dots + a_{mn} x_n \quad (\le, =, \ge b_m)$

$x_1, x_2, \dots x_n > 0$

→ Procedure

1) Check if objective func^n is max z or min z
   If Max (z), leave as it is
   If Min (z), convert as Max (−z)
2) Check whether all 'b_i' are +ve,
   If −ve, multiply by −1 so inequality changes & becomes +ve
3) Express the problem in standard form by introducing a slack or sulphrus variable to make the constraint into equality format
4) Find an initial basic feasible solution, there are m equations & n unknown assign 0 values to n−m variables for finding the solution
   If all the $s_i > 0$, the basic solution is feasible & non-degenerate
   If 1 or more $s_i = 0$, then solution is called degenerate
5) Basics denotes basic variables, B column denotes values of the basic variables while the remainining variables are 0. Coefficient of x in constraint is the body matrix & coefficient of s is the unitary matrix
   C_B column denotes the coefficient of objective functions of the basic variables
6) If all C_j are −ve, then the initial basic feasible solution is the optimal sol^n
   Even if 1 C_j is +ve, the current sol^n needs to be improved
7) Identify the incoming & outgoing variables
   Identify Max C_j & variable corresponding becomes our incoming
   Divide the elements of b column by elements of Key column & put in θ column
   Choose the row with minimum θ & then replace corresponding basic variable with the outgoing variable which becomes the Key row

Q. Maximize $7x_1 + 6x_2$

Subject to $2x_1 + 4x_2 \leq 16$

$3x_1 + 2x_2 \leq 12$

$x_1, x_2 \geq 0$

https://youtu.be/9YKLXFqCy6E?si=tho_HbqNJIBmmwrH

Highly Recommended, same problem is solved here

A. i) Add slack variables

$7x_1 + 6x_2 + 0s_1 + 0s_2$

$2x_1 + 4x_2 + s_1 = 16$

$3x_1 + 2x_2 + s_2 = 12$

ii) Make x_1 & x_2 non-basic & s_1, s_2 basic

$x_1 = 0$ & $x_2 = 0$ $\Rightarrow s_1 = 16$, $s_2 = 12$

iii) Set up initial simplex tableau

	C_j	7	6	0	0		
C_B	basis	x_1	x_2	s_1	s_2	b	θ
0	s_1	2	4	1	0	16	$\frac{16}{2}$
0	s_2	3 (Key element)	2	0	1	12	$\frac{12}{3}$
	$z_j = \sum c_B a_{ij}$	$0\times2+0\times3 = 0$	$0\times4+0\times2 = 0$	$0\times1+0\times0 = 0$	$0\times0+0\times1 = 0$		
	$C_j = c_j - z_j$	7	6	0	0		

Key Column (x_1)

$x_1 = 0$, $x_2 = 0$

$s_1 = 16$, $s_2 = 12$

iv) Iteration 1 $\quad R_2 \to \frac{R_2}{3} \quad R_1 \to R_1 - 2R_2$

	c_j	7	6	0	0		
C_B	basis	x_1	x_2	s_1	s_2	b	θ
0	s_1	0	$\frac{8}{3}$ (Key element)	1	$-\frac{2}{3}$	8	$\frac{8}{8/3} = 3$ ← Key row
7	x_1	1	$\frac{2}{3}$	0	$\frac{1}{3}$	4	$\frac{4}{2/3} = 6$
	z_j	7	$\frac{14}{3}$	0	$\frac{7}{3}$	28	
	C_j	0	$\frac{4}{3}$	0	$-\frac{7}{3}$		

Key Column (x_2)

v) Iteration 2 $\quad R_1 \to R_1 \times \frac{3}{8} \quad R_2 \to R_2 - \frac{2R_1}{3}$

	c_j	7	6	0	0		
C_B	basis	x_1	x_2	s_1	s_2	b	θ
6	x_2	0	1	$\frac{3}{8}$	$-\frac{1}{4}$	3	
7	x_1	1	0	$-\frac{1}{4}$	$1/2$	4	
	z_j	7	6	$1/2$	2	46	
	C_j	0	0	$-1/2$	-2	$\Rightarrow \leq 0$	

Optimal Solution $\Rightarrow$ $x_1 = 4$, $x_2 = 3$, $s_1 = 0$, $s_2 = 0$,

$Z = 46$

Q.

Machine	Time per unit (min) Product A	Product	Product	Machine Capacity (minutes/ day)
M_1	2	3	2	440
M_2	4	–	3	470
M_3	2	5	–	430

Profits per product A, B, C is ₹4, ₹3, ₹6 respectively

Determine daily no. of units to be manufactured for each product

A. Objective Function $\Rightarrow 4x_1 + 3x_2 + 6x_3 + 0S_1 + 0S_2 + 0S_3 = \text{Max } Z$

Subject to
$$2x_1 + 3x_2 + 2x_3 + S_1 = 440$$
$$4x_1 + 0x_2 + 3x_3 + S_2 = 470$$
$$2x_1 + 5x_2 + 0x_3 + S_3 = 430$$

Basic Variables $\Rightarrow S_1 = 440,\ S_2 = 470,\ S_3 = 430$

Non-basic Variables $\Rightarrow x_1 = x_2 = x_3 = 0$

Objective f^n

C_B	C_j	4	3	6	0	0	0		
	Basis	x_1	x_2	x_3	s_1	s_2	s_3	b	θ
0	s_1	a_{11} 2	a_{12} 3	a_{13} 2	1	0	0	440	$\frac{440}{2}$
0	s_2	a_{21} 4	a_{22} 0	a_{23} 3 $\rightarrow$ Key Element	0	1	0	470	$\frac{470}{3}$ $\rightarrow$ Key Row
0	s_3	a_{31} 2	a_{32} 5	a_{33} 0	0	0	1	430	$\frac{430}{0} = \infty$
	$z_j = \Sigma c_B a_{ij}$	0	0	0	0	0	0		
	$C_j = c_j - z_j$	4	3	6	0	0	0		
				$\downarrow$ Key Column					

$R_1 \rightarrow R_1 - \frac{2R_2}{3}$ & $R_2 \rightarrow \frac{R_2}{3}$

x_3 : incoming variable

S_2 : outgoing variable

C_B	C_j	4	3	6	0	0	0		
	Basics	x_1	x_2	x_3	s_1	s_2	s_3	b	θ
0	s_1	$-2/3$	3 $\rightarrow$ Key Element	0	1	$-2/3$	0	$\frac{380}{3}$	$\frac{380}{9}$ $\rightarrow$ Key Row
6	x_3	$4/3$	0	1	0	$1/3$	0	$\frac{470}{3}$	∞
0	s_3	2	5	0	0	0	1	430	86
	$z_j = \Sigma c_B a_{ij}$	8	0	6	0	2	0	940	
	$C_j = c_j - z_j$	-4	3	0	0	-2	0		
			$\uparrow$ Key Column						

x_2 : incoming variable

s_1 : outgoing variable

C_B	C_j / Basics	4 x_1	3 x_2	6 x_3	0 s_1	0 s_2	0 s_3	b	θ
3	x_2	$-2/9$	1	0	$1/3$	$-2/9$	0	$380/9$	
6	x_3	$4/3$	0	1	0	$1/3$	0	$470/3$	
0	s_3	$28/9$	0	0	$-5/3$	$10/9$	0	$1970/9$	
	$z_j = \Sigma c_B a_{ij}$	$22/3$	3	6	1	$4/3$	0	$3200/3$	
	$C_j = c_j - z_j$	$-10/3$	0	0	-1	$-4/3$	0		

Since all $C_j \leq 0$, optimal sol^n is $x_1 = 0$, $x_2 = \frac{380}{9}$, $x_3 = \frac{470}{3}$

$Z_{max} = \frac{3200}{3} = 1066.67$ ₹

Q. Objective F^n is Maximized $\Rightarrow$ Max $Z = 90x_1 + 120x_2$

Constraints $\Rightarrow$ Subject to $x_1 \leq 40$ (ha of pine)

$x_2 \leq 50$ (ha of eucalypt)

$2x_1 + 3x_2 \leq 180$ (days of work)

$x_1 \geq 0$, $x_2 \geq 0$

A. Objective Function $\Rightarrow$ $90x_1 + 120x_2 + 0s_1 + 0s_2 = \text{Max } Z$

Subject to $x_1 + 0x_2 + s_1 = 40$

$0x_1 + x_2 + s_2 = 50$

$2x_1 + 3x_2 + s_3 = 180$

C_b	C_j / Basis	90 x_1	120 x_2	0 s_1	0 s_2	0 s_3	b	θ	
0	s_1	1	0	1	0	0	40	$\frac{40}{0}$	
0	s_2	0	1 Key element	0	1	0	50	50	← Key Column
0	s_3	2	3	0	0	1	180	60	
	z_j	0	0	0	0	0	0		
	C_j	90	120 ↑ Key Row	0	0	0			

$R_2 \to R_2$, $R_1 \to R_1$, $R_3 \to R_1 - 3R_2$

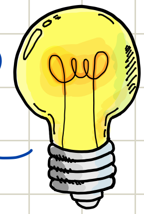
C_b	C_j / Basis	90 x_1	120 x_2	0 s_1	0 s_2	0 s_3	b	θ	
0	s_1	1	0	1	0	0	40	40	
120	x_2	0	1	0	1	0	50	∞	
0	s_3	2	0	0	-3	1	30	15	← Key Column
	z_j	0	120	0	120	0	6000		
	C_j	90 ↑ Key Row	0	0	-120	0			

$R_3 \to \frac{R_3}{2}$, $R_2 \to R_2$, $R_1 \to R_1 - R_3$

	C_j	90	120	0	0	0		
	Basis	x_1	x_2	s_1	s_2	s_3	b	θ
0	s_1	0	0	1	1.5	-0.5	25	16.67
120	x_2	0	1	0	1	0	50	50
90	x_1	1	0	0	-1.5	0.5	15	-10
	z_j	90	120	0	-15	45	7350	
	C_j	0	0	0	15	-45		

Key Element

Key column (consider only +ve ratios)



Key Row

$R_1 \to \frac{R_1}{1.5}$, $R_2 \to R_2 - R_1$, $R_3 \to R_3 + 1.5R_1$

	C_j	90	120	0	0	0	
	Basis	x_1	x_2	s_1	s_2	s_3	b
0	s_2	0	0	0.67	1	-0.33	16.67
120	x_2	0	1	-0.67	0	0.33	33.33
90	x_1	1	0	1	0	0	40
	z_j	90	120	10	0	40	7600
	C_j	0	0	-10	0	-40	≤ 0

$x_1 = 40$, $x_2 = 33.33$

$s_1 = s_2 = s_3 = 0$

$Z = 7600$

Q. Min $Z = x_1 - 3x_2 + 3x_3$

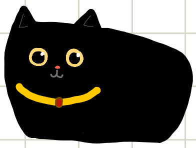
Subject to $3x_1 - x_2 + 2x_3 \leq 7$

$2x_1 + 4x_2 \geq -12$

$-4x_1 + 3x_2 + 8x_3 \leq 10$

$x_1, x_2, x_3 \geq 0$

Solve on Own



Discrete Variable Optimization

→ Design variables are to be selected from set of discrete values

ex: Let the design variable be beam size & 20 beam sizes are available

→ In continuous optimization, objective functions are non-convex & have saddle points and traps the optimization algorithm before reaching intended destination

But discrete setting doesn't have gradients, there is no specific 'path' that can be followed you can only jump b/w seperate points, and not lot of trial & error

→ Branch & Bound Algorithm

Splitting the problem into sub-problems

Estimating the O.F which is used to cut the tree

Q. Maximize $Z = 3x_1 + 5x_2$

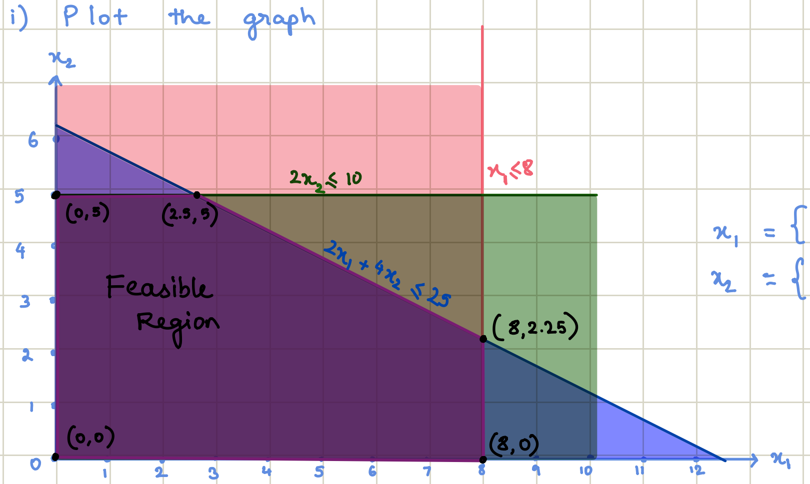
such that $2x_1 + 4x_2 \leq 25$

$x_1 \leq 8$, $2x_2 \leq 10$

x is design variable & can only take +ve integer values

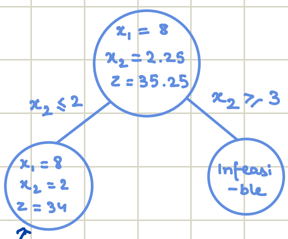
Determine x, A, B, C

A. i) Plot the graph



$x_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

$x_2 = \{0, 1, 2, 3, 4, 5\}$



Final Answer

ii) Find Optimal solution by relaxing integer condition.

$x_1 = 8$, $x_2 = 2.25$, $z = 3(8) + 5(2.25) = 35.25$

iii) Branch out, Keep one of the D.V to be an integer & other D.Vs to take continuous values

Let it partition into $x_1 = 8, x_2 <= 2$ and $x_1 = 8, x_2 >= 3$

↳ $z = 34$ ↳ $z =$ infeasible

So we don't branch out further

Q. Maximize $z = 5x_1 + 4x_2$

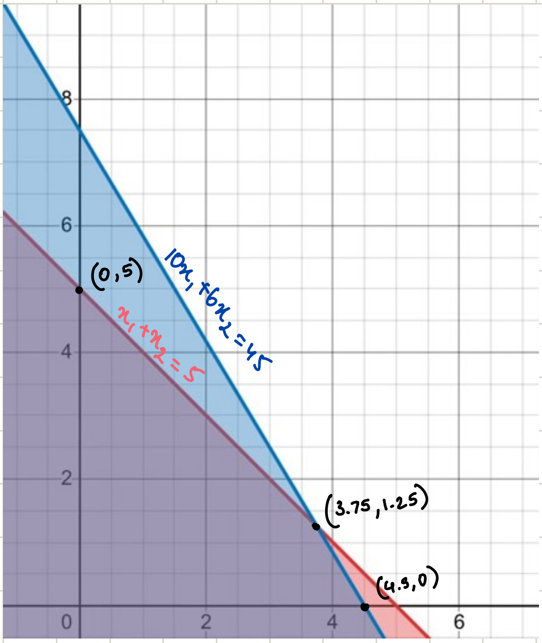
Such that $x_1 + x_2 \leq 5$

$10x_1 + 6x_2 \leq 45$

x is design variable & can only take +ve integer values

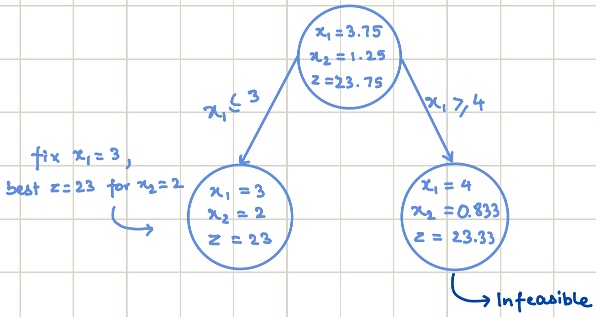
Determine x, A, B, C

A.



Feasible Solution = $(3.75, 1.25)$

↳ $z = 23.75$



Q. Minimize $z = 5x_1 + 4x_2$

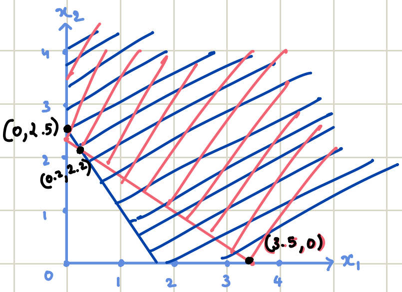
such that $3x_1 + 2x_2 \geq 5$

$2x_1 + 3x_2 \geq 7$

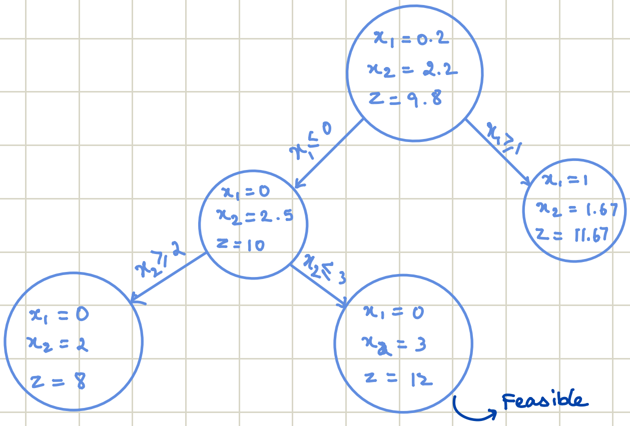
x is design variable & can only take +ve integer values

Determine x, A, B, C

A.



Feasible solution ⇒ $x_1 = 0.2$, $x_2 = 2.2$, $z = 9.8$



→ When to stop branching?

i) When integer solution is obtained

ii) When infeasible solution is obtained

iii) When further branching doesn't improve z

Non-linear Optimisation

→ Optimization involving non-linearities

→ Many NLPs don't have constraints & are called unconstrained NLPs

Q. Maximise $f(x) = -x^2 + 9x + 4$

A. No constraint is given,

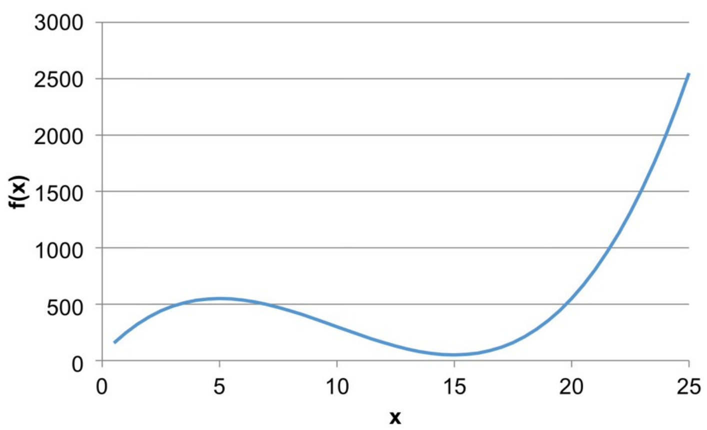
Can be solved using calculus

$f'(x) = -2x + 9 = 0 \Rightarrow x = 4.5$

$f''(x) = -2 \Rightarrow -ve \Rightarrow$ Maxima at $x = 4.5$

So, $Z = -(4.5)^2 + 9(4.5) + 4 = 24.25$

Q. Maximize $f(x) = x^3 - 30x^2 + 225x + 50$



A. Problem is unbounded, no solution point

Q. The Hickory Cabinet and Furniture Company has decided to concentrate on the production of chairs. The fixed cost per month of making chairs is $7,500, and the variable cost per chair is $40. Price is related to demand according to the following linear equation:

$$d = 400 - 1.2p,$$

where d is the demand and p is the price. Develop the nonlinear profit function for this company and determine the price that will maximize profit, the optimal volume, and the maximum profit per month.

A. Profit = Revenue − Cost

Revenue = Units sold × Price = Demand × Price $= d \times p$

$= (400 - 1.2p)(p) = 400p - 1.2p^2$

Cost $= 7500 + 40d$

$= 7500 + 40(400 - 1.2p)$

$= 23500 - 48p$

Profit $= 400p - 1.2p^2 - 48p + 23500$

$= 352p - 1.2p^2 + 23500$

Max profit $\Rightarrow \frac{d(\text{Profit})}{dp} = 352 - 2.4p = 0 \Rightarrow p = 146.67 \Rightarrow$ Max profit $= 49313.33$

↳ Price at which it gives max profit

Optimal Volume $= d = 400 - 1.2p = \underline{\underline{224}}$

Lagrange's Multipliers

→ Transform non-linear optimisation problem into unconstrained form using extra (auxiliary) variable or Lagrange's multiplier for each constraint

$g(x,y) = 0 \Rightarrow$ constraint

$f(x,y) \Rightarrow$ Objective f^n

$\lambda \Rightarrow$ Lagrange's multiplier

$L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$)

→ Lagrange's Multipliers help to solve the problem when the constraints have '=' sign

i) Define the lagrange O.F

ii) Set all parallel derivatives to '0'

iii) Find point of intersection

Q. Maximise $f(x,y) = 4x + 3y$, subject to constraint $x^2+y^2 = 25$

A. $L(x,y,\lambda) = 4x+3y - \lambda(x^2+y^2-25)$

$\frac{\partial L}{\partial x} = 4 - 2\lambda x = 0 \Rightarrow \lambda x = 2$

$\frac{\partial L}{\partial y} = 3 - 2\lambda y = 0 \Rightarrow \lambda y = \frac{3}{2}$

$\frac{\partial L}{\partial \lambda} = x^2+y^2-25 = 0 \Rightarrow x^2+y^2 = 25$

$\left(\frac{2}{\lambda}\right)^2 + \left(\frac{3}{2\lambda}\right)^2 = 25$

$\frac{4}{\lambda^2} + \frac{9}{4\lambda^2} = 25$

$\frac{1}{\lambda^2}\left(\frac{16+9}{4}\right) = 25 \Rightarrow \lambda = \pm\frac{1}{2}, \; x = \pm 4, \; y = \mp 3$

Max occurs at $(4,3)$ & $(-4,-3)$

Value of f^n at these points $= 25$

Q. Maximize $P(x,y) = 50x + 100y$ subject to $x+2y = 100$

A. $L(x,y,\lambda) = (50x+100y) + \lambda(x+2y-100)$

$\frac{dL}{dx} = 50 + \lambda = 0 \Rightarrow \lambda = -50$

$\frac{dL}{dy} = 100 + 2\lambda = 0 \Rightarrow \lambda = -50$

$\frac{dL}{d\lambda} = x+2y-100 = 0 \Rightarrow x+2y = 100$

Then $50x + 100y = 50(x+2y) = 50 \times 100 = 5000$ max

Karush Kuhn Tucker Condition

→ Help solving non linear optimization problems with inequalities

So it is a generalised version of Lagrange's multipliers

→ KKT conditions :

i) Stationarity : $\nabla f(x) - \lambda \nabla g(x) = 0$

ii) Primal Feasibility : Constraints must hold $g(x) \leq 0$

iii) Dual Feasibility : $\lambda \geq 0$ for inequality constraints

iv) Complementary Slackness : $\lambda g(x) = 0$

$L = f - \sum \lambda_i g_i$

Case i) Max $f(x)$

S.T $g_i(x) \leq b_i$

Case ii) Max $f(x)$

ST $g_i(x) \geq b_i$

Case iii) Min $f(x)$

S.T $g_i(x) \leq b_i$

Case iv) Min $f(x)$

ST $g_i(x) \geq b_i$

If $f(x)$ is convex f^n, then H is a +ve semi-definite

If $f(x)$ is concave f^n, then H is a -ve semi-definite

Q. $3.6x_1 - 0.4x_1^2 + 16x_2 - 0.2x_2^2$, Maximise

S.T $2x_1 + x_2 \leq 10$

$x_1, x_2 \geq 0$

A. $L = 3.6x_1 - 0.4x_1^2 + 1.6x_2 - 0.2x_2^2 - \lambda(2x_1 + x_2 - 10)$

$\frac{\partial L}{\partial x} = 3.6 - 0.8x_1 - 2\lambda = 0$

$\frac{\partial L}{\partial y} = 1.6 - 0.4x_2 - \lambda = 0$

$\lambda(2x_1 + x_2 - 10) = 0$

$2x_1 + x_2 - 10 \leq 0$

$x_1, x_2 \geq 0$, $\lambda \geq 0$

Case i) $\lambda = 0$

$3.6 = 0.8x_1 \Rightarrow x_1 = 4.5$

$1.6 = 0.4x_2 \Rightarrow x_2 = 4$

$f(4.5, 4) = 68.9$ ✓

Case ii) $\lambda \neq 0$

$2x_1 + x_2 = 10$

$5.2 - 0.4(2x_1 + x_2) - 3\lambda = 0 \Rightarrow \lambda = \frac{1.2}{3} = 0.4$

$x_1 = 3.5$, $x_2 = 3$, $f(3.5, 3) = 53.9$ ✗

Q. Minimize $f(x,y) = x^2 + y^2$ ST $x + y \leq 1$ and $x, y \geq 0$

A. Set up Lagrangian

$L(x, y, \lambda) = x^2 + y^2 - \lambda(x + y - 1)$

$\frac{\partial L}{\partial x} = 2x - \lambda = 0$ $\qquad$ $\frac{\partial L}{\partial y} = 2y - \lambda = 0$

$\lambda(x + y - 1) = 0$ $\qquad$ $x + y - 1 \leq 0$ $\qquad$ $x, y \geq 0$ $\qquad$ $\lambda \leq 0$

Case i) $\lambda = 0$

$2x = \lambda = 0 \Rightarrow x = 0$

$2y = \lambda = 0 \Rightarrow y = 0$

$f(0, 0) = 0 \Rightarrow$ invalid

Case ii) $\lambda \neq 0$

$\lambda(x + y - 1) = 0$

$x + y - 1 = 0 \Rightarrow x + y = 1$

$2x - \lambda + 2y - \lambda = 0 \Rightarrow x + y = \lambda \Rightarrow \lambda = 1$

$x = 0.5$, $y = 0.5 \Rightarrow f(0.5, 0.5) = 0.5^2 + 0.5^2 = 0.5$

Unconstrained Optimisation

Convex Function

→ a single variable f^n is complex if line segment joining 2 arbitrary points on the function / curve lies above or the curve b/w the points

So,

for all $0 \leq t \leq 1$ & $x_1, x_2 \in X$ where X is domain of function $f(x)$

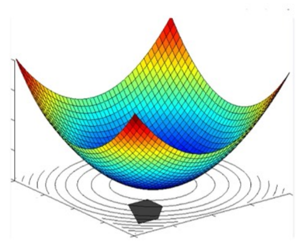
Then,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

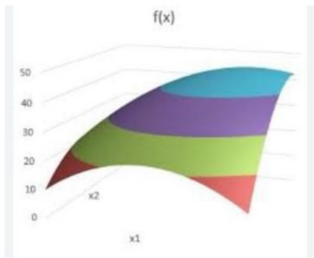
But if

$$f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$$, it is concave function

→ 2 variable convex & concave functions



Convex Function



Concave Function

→ Convex Set is called so if every line segment joining a pair of points, Then set lines lies in the set

For all $0 \leq t \leq 1$ & $x_1, x_2 \in S$

then, $tx_1 + (1-t)x_2 \in S$

→ First order derivative of single variable function

→ For single variable f^n, it represents slope while second order derivative, represents rate of change of slope

→ So, $f'(x) = 0$ at stationary points

$f'(x) = 0$ & $f''(x) < 0$ at local minima

$f'(x) = 0$ & $f''(x) > 0$ at local maxima

→ For multi-variable function, $f'(x)$ is a vector and is called gradient

→ $\nabla f|_{(a,b)} \cdot \hat{v}$ denotes slope of $f(.)$ along the vector $\vec{v}$ about (a,b)

Only holds good if $f(.)$ is continuous

Q. $f(x_1, x_2) = -2x_1 + x_2 + 2x_1^2 + 3x_1x_2 + x_2^2$. Determine directional derivative along $\vec{v} = \frac{\sqrt{3}}{2}\hat{x}_1 + \frac{1}{2}\hat{x}_2$ at $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$

A. $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} -2 + 4x_1 + 3x_2 \\ 1 + 3x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} \text{Slope along } x\text{-axis} \\ \text{Slope along } y\text{-axis} \end{bmatrix}$

$$\nabla f_{(2,-2)} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Directional derivative $= \begin{bmatrix} 0 \\ 3 \end{bmatrix} \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix} = 0 + \frac{3}{2} = \frac{3}{2}$

Stationary Points of Multi-Variable Function

→ 3 Stationary Points :

i) Local Minima ⇒ $(\hat{v})^T H (\hat{v}) > 0$

ii) Local Maxima ⇒ $(\hat{v})^T H (\hat{v}) < 0$

iii) Saddle Point

→ At stationary point, slope = 0 ⇒ Directional derivative = 0 ⇒ Gradient = 0

Q. Find stationary points for $f(x_1, x_2) = -2x_1 + x_2 + 2x_1^2 + 3x_1 x_2 + x_2^2$

A. $\frac{\partial f}{\partial x_1} = -2 + 4x_1 + 3x_2 = 0 \longrightarrow (1)$

$\frac{\partial f}{\partial x_2} = 1 + 3x_1 + 2x_2 = 0 \longrightarrow (2)$

Solving (1) & (2) ⇒ $(x_1, x_2) = (-7, 10)$

Second Order Derivative of Multi-variable function

→ It is a matrix for MVF and is called Hessian matrix

→ It signifies rate of change of slope & depends on location of point of interest

→ Hessian matrix

$$\vec{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Q. Find Hessian matrix for

i) $f(x_1, x_2) = -2x_1 + x_2 + 2x_1^2 + 3x_1 x_2 + x_2^2$

ii) $f(x_1, x_2) = 2x_1^{0.5} + 3\ln(x_2)$

iii) $f(x, y) = x^4 - 32x^2 + y^4 - 18y^2$

iv) $f(x, y) = x^3 + 3xy + y^3$

A. i) $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$

ii) $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -0.5x_1^{-1.5} & 0 \\ 0 & -3x_2^{-2} \end{bmatrix}$

iii) $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 12x^2 - 64 & 0 \\ 0 & 12y^2 - 36 \end{bmatrix}$

iv) $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 6x & 3 \\ 3 & 6y \end{bmatrix}$

→ Square matrix A is +ve definite if $(\vec{x})^T A (x) > 0$

Square matrix A is −ve definite if $(\vec{x})^T A (x) < 0$

So, to find if point is local minima, Hessian matrix must be positive definite

to find if point is local maxima, Hessian matrix must be negative definite

→ Test for +ve definiteness : all of its principal minors are +ve

→ Test for −ve definiteness : its principal minors are alternatively −ve and +ve (1st is −ve)

→ Test for saddle point : if neither +ve or −ve definite

Q. Find if +ve, −ve definite or saddle point

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

ii) $f(x_1,x_2) = 4x_1 + 2x_2 + x_1^2 - 4x_1x_2 + x_2^2$

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

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

v) $f(x,y) = 2x^{0.5} + 3\ln y$

A. i) $H = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$

1st principal minor $= 2$
2nd principal minor $= -4$ ⇒ Saddle point | Solving $2x = 0$, $-2y = 0$ ⇒ $x = 0, y = 0$ is saddle point

ii) $H = \begin{bmatrix} 2 & -4 \\ -4 & 2 \end{bmatrix}$

1st principal minor $= 2$
2nd principal minor $= 4 - 16 = -12$ ⇒ Saddle point | Solving $4 + 2x_1 - 4x_2 = 0$, $2 - 4x_1 + 2x_2 = 0$ ⇒ $x_1 = \frac{4}{3}$, $x_2 = \frac{5}{3}$ is saddle point

iii) $H = \begin{bmatrix} 8 & -1 \\ -1 & 4 \end{bmatrix}$

1st principal minor $= 8$
2nd principal minor $= 32 - 1 = 31$ ⇒ +ve definite | Solving $1 + 8x - y = 0$, $2 - x + 4y = 0$ ⇒ $x = \frac{-6}{31}$, $y = \frac{-17}{31}$ is stationary point

iv) $H = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$

1st principal minor $= 4$
2nd principal minor $= 8 - 9 = -1$ ⇒ Saddle point | Solving $-2 + 4x_1 + 3x_2 = 0$, $1 + 3x_1 + 2x_2 = 0$ ⇒ $x = -7$, $y = 10$ is saddle point

v) $H = \begin{bmatrix} -0.5x^{-1.5} & 0 \\ 0 & \frac{-3}{y^2} \end{bmatrix}$

1st principal minor $= -0.5x^{-1.5}$
2nd principal minor $= \frac{1.5}{x^{1.5}y^2}$ ⇒ −ve definite | Solving $x^{-0.5} = 0$, $\frac{3}{y} = 0$ ⇒ saddle point at $x = 0, y = 0$

Gradient

→ Gradient measures how much the output of a function changes if input is changed a little bit

Gradient Descent

→ Most basic & first-order optimization algorithm which is dependant on first-order derivative of a loss function

→ $w_{new} = w_{old} - \eta \frac{\partial L}{\partial w}$

$w_n = w_{n-1} - \eta \frac{\partial L}{\partial w}$

→ The steps gradient descent takes into the direction of local minimum are determined by learning rate which helps to figure out how fast or slow we move to optimal weights

→ Steps to solving

i) Initialization - Begin with initial guess for parameters which can be random

ii) Compute the gradient - Calculate gradient of objective f^n wrt parameters

iii) Update parameters - Update the parameters using the formula

iv) Repeat - Continue updating parameters until convergence criteria are met, such as small change in the objective function value or parameters

Q. Find local minima of the function $y = (x+5)^2$ starting from the point $x = 3$

A. It reaches minimum value at $y = 0 \Rightarrow x = -5$

$x = -5$ is local & global minima of the function

$X_0 = 3$

Learning rate $= 0.01$

$\frac{dy}{dx} = \frac{d}{dx}(x+5)^2 = 2(x+5)$

$x_1 = x_0 - (\text{learning rate})\left(\frac{dy}{dx}\right)$

$= 3 - 0.01\left(2 \times (3+5)\right)$

$= 3 - 0.16$

$= 2.84$

→ Non-linear optimization can face several challenges like local minima, saddle points and slow convergence