OPERATIONS RESEARCH

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

SEM-1

OR-1

Operational Research

A problem requiring maximization or minimization of a numerical function of several decision variables satisfying a number of constraints is defined as <u>Optimization problem</u>.

- <u>Construction of an L.P.P. Problem:-</u>

Reddy Mikks produces both interior & exterior paints from two raw materials, M1 and M2.

	Tons of raw material per ton of		Maximum daily Availability (tons)
	Exterior paint	Interior paint	
Raw material, M1	6	4	24
Raw material, M2	1	2	6
Profit per ton ($1000)	5	4	

Reddy Mikks wants to determine the optimum (best) product mix of interior & exterior paints that maximizes the total daily profit.

So, we need to determine the daily amounts to be produced of exterior & interior paints. Thus the variables of the model are defined as

x_1 = Tons produced daily of exterior paint

x_2 = Tons produced daily of interior paint

Profit from exterior paint = $5x_1$ (thousand) dollars

" " interior " = $4x_2$ (") "

Let Z represents the maximum profit, so the problem is:

$$\text{Maximize } Z = 5x_1 + 4x_2$$

Constraints are

$$6x_1 + 4x_2 \leq 24 \quad \text{(Raw material M1)}$$

$$x_1 + 2x_2 \leq 6 \quad \text{(Raw material M2)}$$

$$\text{and } x_1 \geq 0, \; x_2 \geq 0$$

Any values of x_1, x_2 that satisfies all three constrains constitute a FEASIBLE SOLUTION. <u>eg</u>:- x_1 = 3tons/day ; x_2 = 1 ton/day.

The goal is to find out the best feasible solution, i.e., the OPTIMUM SOLUTION that maximize the total profit Z.

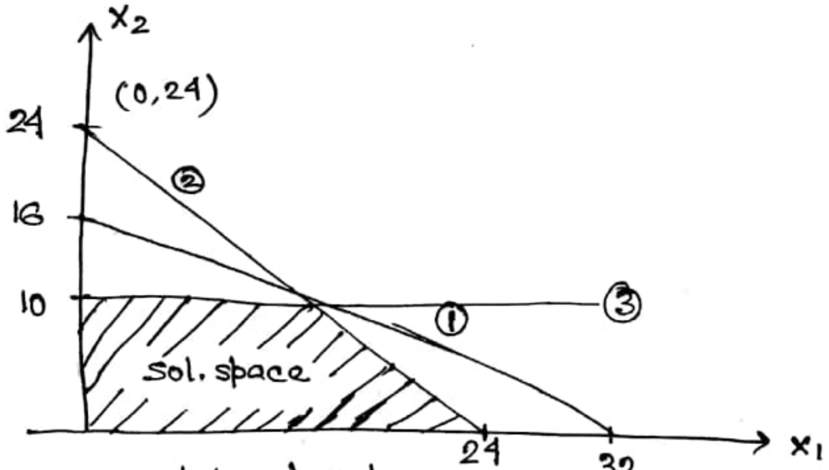
- **Graphical LP Solution :-** The graphical solution includes two steps:
  1. Determination of the feasible solution space.
  2. Determination of the optimum solution from among all points in the solution space.

Ex :- Two types of TV sets to be manufactured Black & white and colours. Total TV chases available = 24 units.
Total production hours = 160 hours
Total colour tube = 10 units.
B&W requires 5 hours of assembling.
Colour set " 10 " "
Profit for BW Rs. 60/unit. Profit for colour Rs. 150/unit.

Solution:- Total no. of black & white TV manufactured : x_1
" " " colour TV " : x_2

Max $Z = 60x_1 + 150x_2$, subject to the constraints

$$5x_1 + 10x_2 \le 160 \quad \text{--- (1)}$$
$$x_1 + x_2 \le 24 \quad \text{--- (2)}$$
$$x_2 \le 10 \quad \text{--- (3)}$$
$$x_1, x_2 \ge 0$$



Optimum solution points are: (0,0), (0,10), (24,0), (12,10), (16,8)

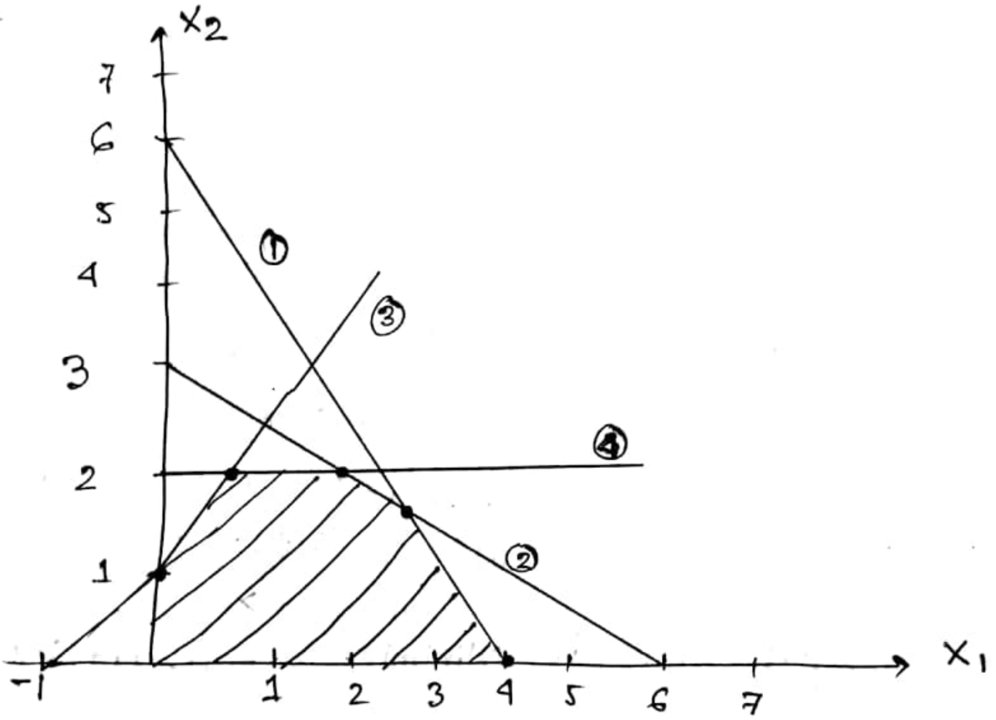
Now, Z = 0 at (0,0)
= 1500 at (0,10)
= 1440 at (24,0)
= 2220 at (12,10)
= 2160 at (16,8)

So, Max profit = 2220 Rs.

Ex.2. $\quad \text{Max } Z = 5x_1 + 4x_2$

subject to
$$6x_1 + 4x_2 \le 24$$
$$x_1 + 2x_2 \le 6$$
$$-x_1 + x_2 \le 1$$
$$x_2 \le 2$$
$$x_1, x_2 \ge 0$$

Solution:-



Solution points are: $(1,2)$, $(0,1)$, $(2,2)$, $(3,1.5)$, $(4,0)$

$Z_{(1,2)} = 13$; $Z_{(0,1)} = 4$; $Z_{(2,2)} = 18$, $Z_{(3,1.5)} = 21$,

$Z_{(4,0)} = 20$.

The solution is $Z = 5 \times 3 + 4 \times 1.5 = 21$, i.e., $\max Z = 21$.

Convex Line:- It is a collection of points denoted as R in which a line joining any two points will lie in R itself.

Extreme Point:- A point E in the convex set is called Extreme point if it is not possible to locate 2 dist so that line segment through them will contain E.

Note:- If minimum or maximum of a linear function defined over feasible region exist, then it must be at one of the extreme point.

Ex.3. Ozark farms uses at least 800 lb of special feed daily. The special feed is a mixture of corn and soybean meal with the following compositions:

Feedstuff	lb per lb of feedstuff: Protein	Fiber	Cost ($/lb)
Corn	.09	.02	0.30
Soybean meal	.60	.06	0.90

The dietary requirements of the special feed are at least 30% protein and at most 5% fiber.
The goal is to determine the daily minimum cost feed mix.

Solution:- The decision variables of the model are

x_1 = lb of corn in the daily mix

x_2 = lb of soybean meal in the daily mix

$$\text{Minimize } z = 0.3x_1 + 0.9x_2$$

Constraints are:

$$x_1 + x_2 \geq 800$$

$$0.09x_1 + 0.60x_2 \geq 0.3(x_1 + x_2)$$

$$0.02x_1 + 0.06x_2 \leq 0.5(x_1 + x_2)$$

$$x_1, x_2 \geq 0$$

The complete model is Minimize $z = 0.3x_1 + 0.9x_2$

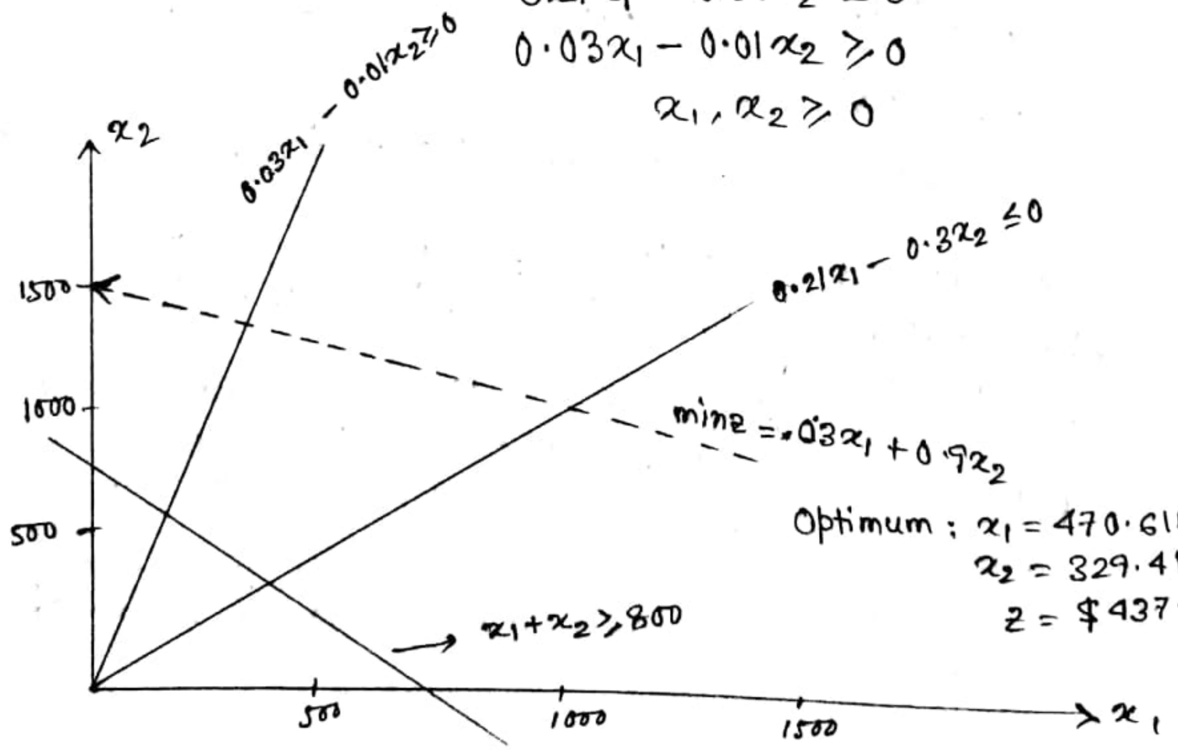
subject to

$$x_1 + x_2 \geq 800$$

$$0.21x_1 - 0.30x_2 \leq 0$$

$$0.03x_1 - 0.01x_2 \geq 0$$

$$x_1, x_2 \geq 0$$



Optimum: $x_1 = 470.6$ lb

$x_2 = 329.4$ lb

$z = \$437.64$

General Programming Problem:-

$$\text{Min or Max } Z = \sum_i \sum_j c_{ij} X_{ij}$$

$$\text{s.t.} \quad \sum \sum a_{ij} x_{ij} \geq \text{ or } \leq \text{ or } = b_j$$

$$\text{All } x_{ij} \geq 0$$

It will be linear if Z is a linear function as well as the constraints are also linear.

NOTES:- ① Shades will be given accordingly as $x_1 = x_2 = 0$ & relative conditions satisfied or not. Suppose it shows $0 \geq 1200$, then

and if it shows $0 \leq 400$, then

② Infeasible solution means no common region.

③ Unbounded means common region is not closed.

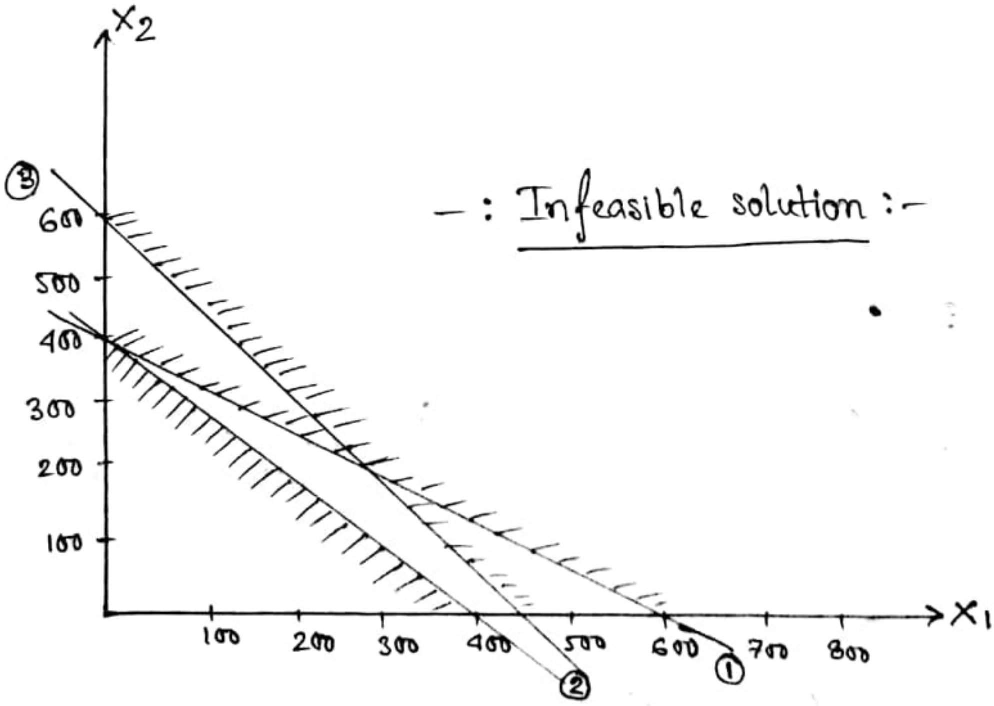
④ Redundancy means if we eliminate some constraints then it will not effect the optimum solution.

Q.1. $\text{Min } Z = 200X_1 + 300X_2$

$$\text{s.t.} \quad 2X_1 + 3X_2 \geq 1200$$
$$X_1 + X_2 \leq 400$$
$$2X_1 + 1.5X_2 \geq 900$$
$$X_1, X_2 \geq 0$$

Solution:-

-: Infeasible solution :-



Q.2. $\text{Min } z = 120x_1 + 160x_2$

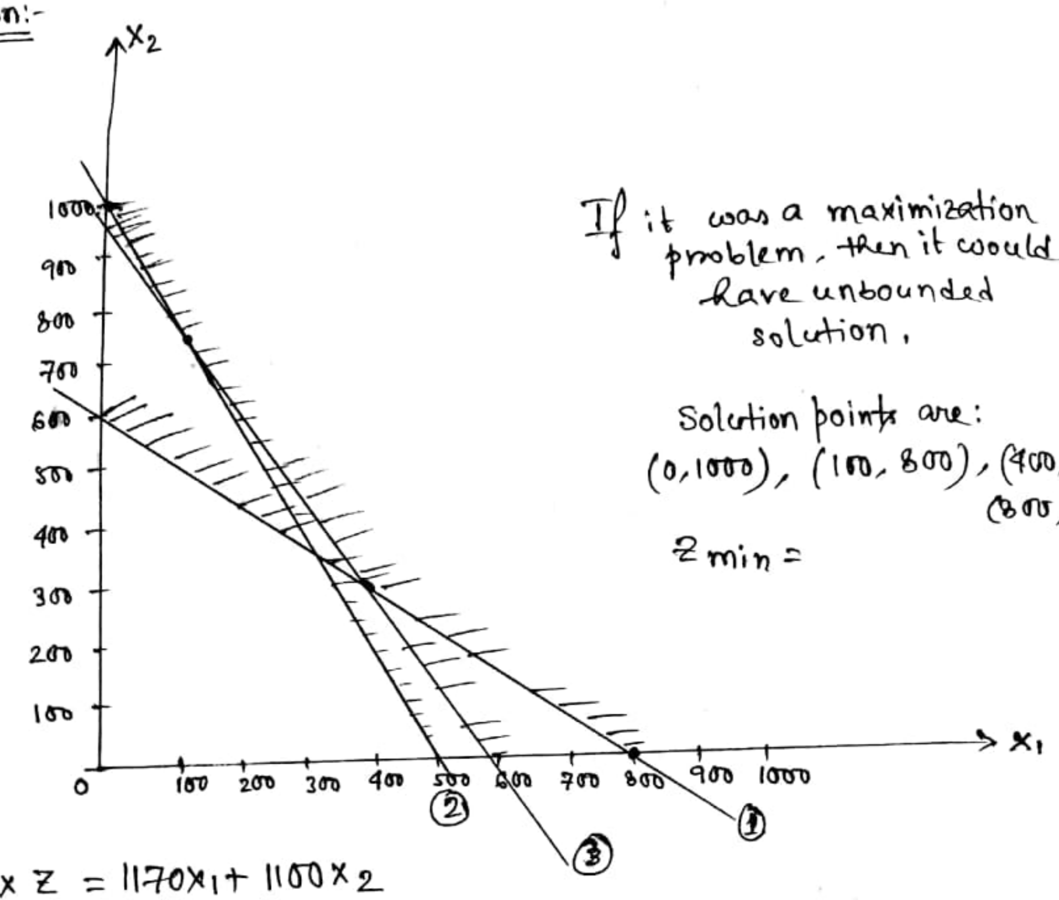
s.t. $3x_1 + 4x_2 \geq 2400$

$2x_1 + x_2 \geq 1000$

$5x_1 + 3x_2 \geq 2900$

$x_1, x_2 \geq 0$

Solution:-



If it was a maximization problem, then it would have unbounded solution.

Solution points are:
(0, 1000), (100, 800), (400, 300), (800, 0)

z min =

Q.3. $\text{Max } Z = 1170x_1 + 1100x_2$

s.t. $9x_1 + 5x_2 \geq 500$

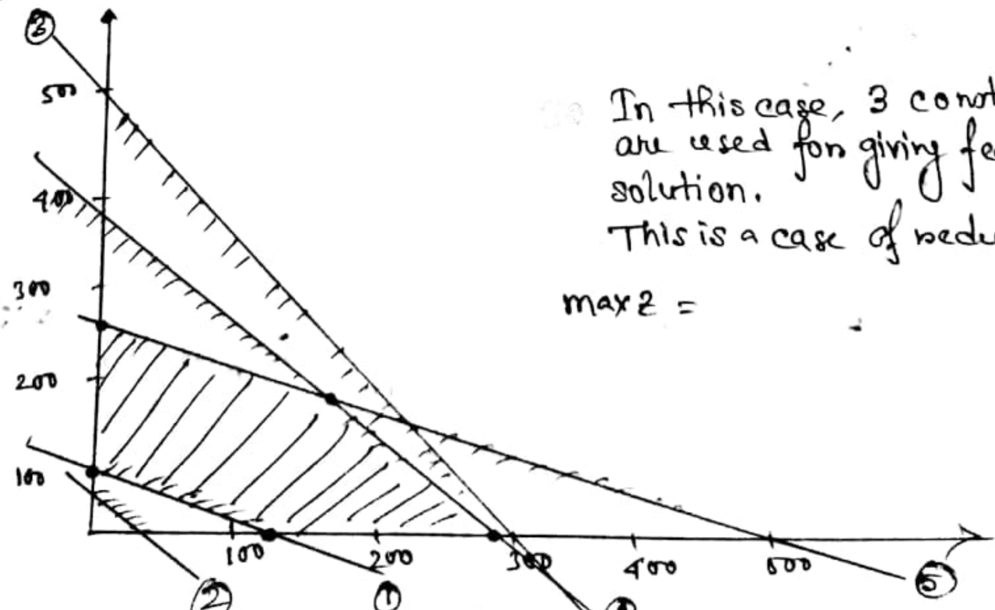
$7x_1 + 9x_2 \geq 300$

$5x_1 + 3x_2 \leq 1500$

$7x_1 + 5x_2 \leq 1900$

$2x_1 + 4x_2 \leq 1000$

Solution:-



In this case, 3 constraints are used for giving feasible solution.
This is a case of redundant.

max z =

Simplex Method:- Iterative procedure of reaching the optimum solution by starting with area of the basic feasible solution and move to adjoining extreme points and validating the solution for optimality.

- Simplex is a type of convex set.
- Developed by DANTZIG.

Slack variable is a variable introduced when constraints are $\leq$ type. Slack = Deficiencies.

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1$$

Introduce x_{n+1} with cost coefficient 'zero'. Then

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1$$

Objective function $Z = c_1x_1 + \dots + c_nx_n + 0.x_{n+1}$

If RHS is negative, multiply by -1 both sides, and continue simplex method.

Surplus variable is introduced when constraints are $\geq$ type.

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \geq b_1$$

Introduce a surplus variable x_{n+1} with coefficient 'zero'.

$$a_{11}x_1 + \dots + a_{1n}x_n - x_{n+1} = b_1$$

Case of minimization:- Convert the problem into maximization
minimize $Z' =$ maximize$(-Z)$
and follow the usual approach.

Case of Maximization:-
- Examine $z_j - c_j$'s. If all are ≥ 0, stop the operation, optimum solution is reached. Then consider that x_j for which $z_j - c_j$ is the least.
- In case of tie, choose the x_j for which the subscript is lower.
- If $z_j - c_j = 0$ for all basic & non-basic variable, then multiple solution exist for the LPP.
- Consider the ratio, $\left\{ \frac{x_B}{y_{ij}}, y_{ij} > 0 \right\}$
  If all $y_{ij} \leq 0$, stop the iteration. LPP has unbounded solution.
  OW consider the variable for which ratio is minimum.
- In case of tie, break it by either choosing the variable for which y_{ij} is the maximum or from top most row.

Ex:- Solve the following LPP by simplex method:-

$$\text{Max } Z = 2x_1 + 4x_2$$

$$\text{s.t. } 2x_1 + 3x_2 \le 48$$

$$x_1 + 3x_2 \le 42$$

$$x_1 + x_2 \le 21$$

$$x_1, x_2 \ge 0$$

Solution:- Let x_3, x_4, x_5 are slack variables with cost coeff. '0'.

So, our problem is $\max Z = 2x_1 + 4x_2 + 0x_3 + 0x_4 + 0x_5$

$$\text{s.t. } 2x_1 + 3x_2 + x_3 + 0x_4 + 0x_5 = 48$$

$$x_1 + 3x_2 + 0x_3 + x_4 + 0x_5 = 42$$

$$x_1 + x_2 + 0x_3 + 0x_4 + x_5 = 21$$

Identify an initial BFS by letting $x_1, x_2 = 0$; so,

$$x_3 = 48, \ x_4 = 42, \ x_5 = 21$$

	C_j	2	4	0	0	0	
C_B	BV	x_1	x_2	x_3	x_4	x_5	X_B
0	x_3	2	3	1	0	0	48
0	← x_4	1	3	0	1	0	42
0	x_5	1	1	0	0	1	21
	$z_j - c_j$	-2	-4 ↑	0	0	0	$z = 0$

Decision:- If all $z_j - c_j$ are +ve then stop. Current solution is the optimal solution. If not choose the column with least -ve value, that is minimum $z_j - c_j$. Choose the variable as entering variable.

$$\min\{-2, -4\} = -4$$

$\therefore x_2$ enters the basis.

Now, $\min\left\{\frac{48}{3}, \frac{42}{3}, \frac{21}{1}\right\} = 14$

[Determine the ratio, $\frac{X_B}{y_{ij}}$ for $y_{ij} > 0$ for the column corresponding to entering variable]

take the variable corresponding to $\min\left\{\frac{X_B}{y_{ij}}\right\}$; $y_{ij} > 0$

$\therefore x_4$ leaves the basis.

$$R_2' = \frac{R_2}{3}$$

$$R_1' = R_1 - R_2$$

$$R_3' = R_3 - \frac{1}{3} R_2$$

	c_j	2	4	0	0	0	X_B
C_B	BV	X_1	X_2	X_3	X_4	X_5	
0	X_3	$\boxed{1}$	0	1	-1	0	6 →
4	X_2	$\frac{1}{3}$	1	0	$\frac{1}{3}$	0	14
0	X_5	$\frac{2}{3}$	0	0	$-\frac{1}{3}$	1	7
	$Z_j - C_j$	$\frac{4}{3}-2 = \frac{-2}{3}$ ↑	0	0	$4/3$	0	

$\therefore \min\{-2/3, 4/3\} = -2/3$

$\therefore X_1$ enters the basis

$$\min\left\{\frac{6}{1}, \frac{16}{1/3}, \frac{7}{2/3}\right\} = 6$$

$\therefore X_3$ leaves the basis.

$R_1' = R_1$

$R_2' = R_2 - \frac{1}{3}R_1$

$R_3' = R_3 - \frac{2}{3}R_1$

	c_j	2	4	0	0	0	X_B
C_B	BV	X_1	X_2	X_3	X_4	X_5	
2	X_1	1	0	1	-1	0	6
4	X_2	0	1	$-1/3$	$2/3$	0	12
0	X_5	0	0	$-2/3$	$1/3$	1	3
	$Z_j - C_j$	0	0	$2/3$	$2/3$	0	

Since all $Z_j - C_j \geq 0$

$\Rightarrow$ the optimum solution is reached, i.e.

$x_1 = 6, x_2 = 12, x_5 = 3, x_3 = 0, x_4 = 0$

and $\max z = 60$.

Scanned by CamScanner

- **BIG M - Method :-** (By CHARNE'S)

Q. $\text{Min } Z = x_1 + x_2 + x_3$

s.t. $x_1 + 2x_2 \geq 3$

$x_2 + 7x_3 \leq 6$

$x_1 - x_2 + 5x_3 = 5$

$x_j \geq 0 \quad (j = 1, 2, 3)$

Sol. Maximize $z = -x_1 - x_2 - x_3$

Slack variable x_5 and surplus variable x_4.

So, $\text{Max } Z = -x_1 - x_2 - x_3$

s.t. $x_1 + 2x_2 - x_4 = 3$

$x_2 + 7x_3 + x_5 = 6$

$x_1 - x_2 + 5x_3 = 5$

Now add two variables (artificial variable) x_6 and x_7 with cost coefficient very high (for very high cost coeff. we will get soln. in first few iterations)

Now problem reduces to

$$\text{Max } Z = -x_1 - x_2 - x_3 + 0x_4 + 0x_5 - M(x_6 + x_7)$$

s.t. $x_1 + 2x_2 + 0.x_3 - x_4 + 0x_5 + x_6 + 0x_7 = 3$

$0x_1 + x_2 + 7x_3 + 0x_4 + x_5 + 0x_6 + 0x_7 = 6$

$x_1 - 3x_2 + 5x_3 + 0x_4 + 0x_5 + 0x_6 + x_7 = 5$

IBFS,
$x_1 = x_2 = x_3 = 0$
$x_6 = 3$
$x_7 = 5$
$x_5 = 6$

	c_j	-1	-1	-1	0	0	$-M$	$-M$	
C_B	BV	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_B
$-M$	x_6	1	2	0	-1	0	1	0	3
0	x_5	0	1	7	0	1	0	0	6 →
$-M$	x_7	1	-3	5	0	0	0	1	5
$Z_j - C_j$		$-2M+1$	$M+1$	$-5M+1$ ↑	M	0	0	0	

$\text{Min}\{-2M+1, M+1, -5M+1, M\} = -5M+1$

$\therefore x_3$ enters

$\therefore \text{Min}\left\{\frac{6}{7}, \frac{5}{5}\right\} = \frac{6}{7}$

$\therefore x_5$ leaves.

$\therefore$ $R_1' \to R_1$

$R_2' \to R_2/7$

$R_3' \Rightarrow R_3 - \frac{5}{7}R_2$

	c_j	-1	-1	-1	0	0	-M	-M	
C_B	BV	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_B
-M	x_6	1	2	0	-1	0	1	0	3
-1	x_3	0	1	1	0	1/7	0	0	6/7
-M	x_7	1	$-\frac{26}{7}$	0	0	-5/7	0	1	5/7 $\longrightarrow$
$Z_j - C_j$		$-2M+1$	$-2M+\frac{26M}{7}$	0	M	$-\frac{1}{7}+\frac{5M}{7}$	0	0	

$$\text{Min}\left\{-2M+1,\ \frac{12M+6}{7},\ M,\ \frac{5M-1}{7}\right\} = -2M+1$$

$\therefore$ x_1 enters

$$\text{Min}\left\{3/1,\ 5/7\right\} = 5/7$$

$\therefore$ x_7 leaves

$R_1' = R_1 - R_3$

$R_2' = R_2$

$R_3' = R_3$

	c_j	-1	-1	-1	0	0	-M	-M	
C_B	BV	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_B
-M	x_6	0	40/7	0	-1	5/7	1	0	16/7 $\longrightarrow$
-1	x_3	0	1	1	0	1/7	0	0	6/7
-1	x_1	1	-26/7	0	0	-5/7	0	1	5/7
$Z_j - C_j$		0	$\frac{-40M+32}{7}$	0	M	-5/7	M	—	

x_2 enters & x_6 leaves.

Note:- If $Z_j - C_j \geq 0$ but artificial var. is present in BV table, then there will be infeasible solution.

$R_1' = \frac{7}{40} R_1$; $R_2' = R_2 - R_1/40$; $R_3' = R_3 + \frac{26}{7} R_2$

	c_j	-1	-1	-1	0	0	X_B
C_B	BV	x_1	x_2	x_3	x_4	x_5	
-1	x_2	0	1	0	$-7/40$	$1/8$	$2/5$
-1	x_3	0	0	1	$1/40$	$1/8$	$4/5$
-1	x_1	1	0	0	$-13/20$	$-1/4$	$11/5$
	$Z_j - C_j$	0	0	0	$4/5$	0	

$\therefore$ $x_1 = 11/5$, $x_2 = 2/5$, $x_3 = 4/5$ is the optimal solution.

Max $Z^* = -17/5$, $\therefore$ Min $Z = 17/5$.

• TWO PHASE METHOD:-

PHASE - I : Min artificial variables
PHASE - II : Original obj. function (min/max)

Ex:- Min $Z = 6x_1 + 21x_2$

s.t. $x_1 + 2x_2 \geq 3$ $\qquad x_1 + 2x_2 - x_3 + 0x_4 + 0x_5 + x_6$

$x_1 + 4x_2 \geq 4$

$x_1, x_2 \geq 0$

solution:- Max $Z^* = -6x_1 - 21x_2 + 0.x_3 + 0x_4$

s.t. $x_1 + 2x_2 - x_3 + 0x_4 + x_5 + 0x_6 = 3$

$x_1 + 4x_2 + 0x_3 - x_4 + 0x_5 + x_6 = 4$

$x_j \geq 0 \ \forall \ j = 1(1)6$.

Phase I:- Max $Z^* = -(x_5 + x_6)$

s.t. $x_1 + 2x_2 - x_3 + 0x_4 + x_5 + 0x_6 = 3$

$x_1 + 4x_2 + 0x_3 - x_4 + 0x_5 + x_6 = 4$

$x_j \geq 0 \ \forall \ j = 1(1)6$.

	c_j	0	0	0	0	-1	-1	X_B
C_B	BV	x_1	x_2	x_3	x_4	x_5	x_6	
-1	x_5	1	2	-1	0	1	0	3
-1	x_6	1	4	0	-1	0	1	4
	$z_j - c_j$	-2	-6	1	1	0	0	

~~$R_2' = R_2 - 2R_1$~~

~~$R_1' = R_1 - R_2$~~

$R_2' = \frac{R_2}{4}$

$R_1' = R_1 - \frac{R_2}{2}$

	c_j	0	0	0	0	-1	-1	
C_B	BV	x_1	x_2	x_3	x_4	x_5	x_6	X_B
-1	x_5	1/2	0	-1	1/2	1	-1/2	1 →
0	x_2	1/4	1	0	-1/4	0	1/4	1
	$z_j - c_j$	-1/2	0	1	-1/2 ↑	0	1	

$R_1' = 2R_1$

$R_2' = R_2 + \frac{1}{2}R_1$

	c_j	0	0	0	0	
C_B	BV	x_1	x_2	x_3	x_4	X_B
0	x_4	1	0	-2	1	2
0	x_2	$\frac{1}{2}$	1	-1/2	0	3/2
	$z_j - c_j$	0	0	0	0	

<u>Phase II</u> :-

	c_j	-6	-21	0	0	
C_B	BV	x_1	x_2	x_3	x_4	X_B
0	x_4	1	0	-2	1	2 →
-21	x_2	1/2	1	-1/2	0	3/2
	$z_j - c_j$	-9/2 ↑	0	21/2	0	

$R_1' \to R_1$

$R_2' \to R_2 - \frac{1}{2}R_1$

	C_j	-6	-21	0	0	
C_B	BV	x_1	x_2	x_3	x_4	X_B
-6	x_1	1	0	-2	1	2
-2	x_2	0	1	$3/2$	$-1/2$	$1/2$
	$z_j - c_j$	0	0	$3/2$	$9/2$	

$z^* = -45/2, \; x_1 = 2, \; x_2 = 1/2.$

$\therefore \text{Min } z = 45/2.$

DUAL SIMPLEX METHOD

PRIMAL:- $\text{Max } Z = 3x_1 + 5x_2$

$$x_1 \le 4$$
$$2x_2 \le 12$$
$$3x_1 + 2x_2 \le 18$$
$$x_1, x_2 \ge 0$$

$$\begin{cases} x_1 + 0x_2 \le 4 \\ 0x_1 + 2x_2 \le 12 \\ 3x_1 + 2x_2 \le 18 \end{cases}$$

Solution:- **Dual:-** $\text{Min } Z = 4\omega_1 + 12\omega_2 + 18\omega_3$

s.t. $\omega_1 + 3\omega_3 \ge 3$

$2\omega_2 + 2\omega_3 \ge 5$

$\omega_1, \omega_2, \omega_3 \ge 0$

$\omega_1 + 0\omega_2 + 3\omega_3 \ge 3$

$0\omega_1 + 2\omega_2 + 2\omega_3 \ge 5$

Dual simplex Algorithm:-

$$\text{Max } Z = -4\omega_1 - 12\omega_2 - 18\omega_3 + 0.\omega_4 + 0.\omega_5$$

s.t. $-\omega_1 - 3\omega_3 + \omega_4 = -3$

$-2\omega_2 - 2\omega_3 + \omega_5 = -5$

Table 1:-

	C_j	-4	-12	-18	0	0	
C_B	BV	ω_1	ω_2	ω_3	ω_4	ω_5	Solution (b_j)
0	ω_4	-1	0	-3	1	0	-3
0	ω_5	0	-2	-2	0	1	-5 →
	$z_j - c_j$	4	12	18	0	0	

most negative

Existing variable condition:- Min b_j's $\min\{-3,-5\} = -5$

$\Rightarrow \omega_5$ exits.

Entering variable condition:- $\text{Min}\left\{\frac{z_j - c_j}{a_{ij}}, a_{ij} < 0\right\}$ of BV's

$\therefore \min\left\{\frac{12}{2}, \frac{18}{2}\right\} = 6.$

$\Rightarrow \omega_2$ enters.

Table 2:-

$R_1' = R_1, \; R_2' = R_2/-2$

	c_j	-4	-12	-18	0	0	
C_B	BV	ω_1	ω_2	ω_3	ω_4	ω_5	X_B
0	ω_4	-1	0	-3	1	0	-3
-12	ω_2	0	1	1	0	$-1/2$	$5/2$
	$z_j - c_j$	4	0	6	0	6	

ω_4 leaves and ω_3 enters the basis.

$R_1' = R_1/-3, \; R_2' = R_2 + \frac{1}{3}R_1$

Table 3:-

	c_j	-4	-12	-18	0	0	
C_B	BV	ω_1	ω_2	ω_3	ω_4	ω_5	X_B
-18	ω_3	$1/3$	0	1	$-1/3$	0	1
-12	ω_2	$-1/3$	1	0	1	$-1/2$	$3/2$

Since X_B (solutions) ≥ 0, so, $\omega_1 = 0, \omega_2 = 3/2, \omega_3 = 1$ are the optimal solution.

So, $\min z = +36.$

Opt. (primal) = Opt. (Dual).

(NLP)

Affine Combination:- should be linearly independent and $\sum a_i = 1$.

$u = x - y$ is not affine since $1 + (-1) = 0$

$u = 2x - y$ is affine.

$u = \tfrac{1}{2} + \tfrac{1}{2}y$ is convex.

Convex Combination:- Affine combination with non-negative a_i's.

Column Span:-

$$A = \begin{bmatrix} 1 & 0 & 2 \\ & 2 & 3 \\ 1 & & 1 \\ 2 & 1 & \end{bmatrix}$$

$$M(A) = 1\begin{pmatrix}1\\1\\2\end{pmatrix} + 2\begin{pmatrix}0\\2\\1\end{pmatrix} + 3\begin{pmatrix}2\\3\\1\end{pmatrix} = \begin{pmatrix}7\\14\\7\end{pmatrix}$$

$A\underset{\sim}{x} = 0 \longrightarrow$ Null vector of A.

$S = \{x : Ax = 0\}$ called null space.

is subspace since $Ax = 0$, $Ay = 0$

$A(x+y) = 0$

$A(\alpha x) = 0$

Formulation:- 1 Roll $\longrightarrow$ 20 feet, cost $\rightarrow$ 100Rs.

Order: 5 feet – 100 pieces

7 feet – 150 pieces

9 feet – 75 feets

Pattern:-

DV	Pattern	5'	7'	9'
x_1	c_1	4	0	0
x_2	c_2	2	1	0
x_3	c_3	2	0	1
x_4	c_4	1	2	0
x_5	c_5	0	1	1
x_6	c_6	0	0	2

Constraints: $4x_1 + 2x_2 + 2x_3 + x_4 \geq 100$

$x_2 + 4x_4 + x_5 \geq 150$

$x_3 + 2x_5 + x_6 \geq 75$

x_i are non negative integers.

$$\text{Min} \sum_{i=1}^{6} x_i$$

Quadratic Form:-

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$$f(x) = (x_1 \quad x_2)\, A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= (x_1 + 2x_2 \quad 2x_1 + x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= x_1^2 + 2x_1x_2 + 2x_1x_2 + x_2^2$$

To show pd,

(i) eigen values > 0

(ii) $a_{11} > 0$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{vmatrix} > 0$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

$$f(x) = (x_1 \quad x_2 \quad x_3) \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= x_1^2 + 3x_2^2 + 2x_3^2 + 4x_1x_2 + 6x_1x_3 + 2x_2x_3$$

Given $f(x) = x_1^2 - 3x_2^2 + x_3^2 - 2x_1x_3$

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Multivariate Functions & Derivaties:-

$$f(x_1, x_2) = x_1 + x_2$$

$$\nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$f(x) = x_1 + x_2 = 0 + (1 \quad 1) \begin{pmatrix} x_1 - 0 \\ x_2 - 0 \end{pmatrix}$$

$$f(x_1, x_2) = x_1 x_2 \qquad \nabla f(x) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$

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

$$\nabla f(x, y) = \begin{pmatrix} 4x + 3y \\ 3x - 2y \end{pmatrix}$$

$$H_f = \begin{pmatrix} 4 & 3 \\ 3 & -2 \end{pmatrix}$$

$f(x) = x^t A x$

$\nabla f = 2Ax$

$H_f = 2A$

$$g(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= 3x_1^2 - 4x_1x_2 + 4x_2^2$$

$$\nabla g = \begin{pmatrix} 6x_1 - 4x_2 \\ -4x_1 + 8x_2 \end{pmatrix} \qquad H_g = \begin{pmatrix} 6 & -4 \\ -4 & 8 \end{pmatrix} = 2A.$$

Temple Problem:-

Rs	Weight	Quantity
1	147	?
5	89	?
10	181	?
2	146	?

Total Weight = 1927

Min $|\sum w_i x_i - 1927.08|$

s.t. x_i's are non-negative integers.

To show $S = \{(x,y,z) : x + 2y - z = 4\} \subset \mathbb{R}^3$

Let $(x_1, y_1, z_1), (x_2, y_2, z_2)$

$\lambda(x_1, y_1, z_1) + (1-\lambda)(x_2, y_2, z_2)$

$= (\lambda x_1 + \bar{\lambda} x_2, \lambda y_1 + \bar{\lambda} y_2, \lambda z_1 + \bar{\lambda} z_2)$

$= \lambda x_1 + \bar{\lambda} x_2 + 2(\lambda y_1 + \bar{\lambda} y_2) - (\lambda z_1 + \bar{\lambda} z_2)$

$= (\lambda x_1 + 2\lambda y_1 - \lambda z_1) + (\bar{\lambda} x_2 + 2\bar{\lambda} y_2 - \bar{\lambda} z_2)$

$= \lambda 4 + \bar{\lambda} 4 = 4.$

Primal

Min $c^t x$

s.t. $Ax \geq b$

$x \geq 0$

Dual

Max $b^t y$

s.t. $A^t y \leq c$

$y \geq 0$

Min $2x_1 + x_2 - x_3$

s.t. $x_1 - x_2 + 4x_3 \geq 5$

$3x_1 - 5x_2 + x_3 \leq -1$

$x_1 + x_2 + 10x_3 = 5$

$x_1 \geq 0, x_2 \leq 0, x_3 \in \mathbb{R}$

Dual:-

Max $5y_1 + y_2 + 5y_3$

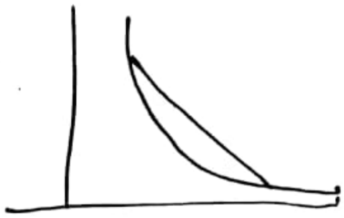
s.t. $y_1 - 3y_2 + y_3 \leq 2$

$y_1 + 5y_2 + y_3 \leq 1$

$4y_1 - y_2 + 10y_3 = -1$

$y_1 \geq 0, y_2 \leq 0, y_3 \in \mathbb{R}$

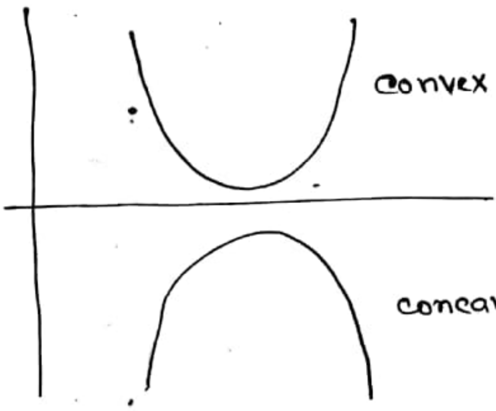
Convexity check :- 1. $f(x) = 1/x$



2. $H(x) = \frac{2}{x^3} > 0 \;\forall\; x > 0$

Second derivitive $> 0 \;\forall\; x > 0$

$H(x)$ is Hessian mtx, need to be psd mtx. to be convex function.



For convex function:-

$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

$\lambda \in (0,1)$, $\forall\; x, y \in S$.

Convex Sets:- A nonempty set S in R^n is said to be convex if the line segment joining any two points in the set is contained in S.

Equivalently, $x, y \in S$, $\lambda \in [0,1]$, the $\lambda x + (1-\lambda) y \in S$.

Example:-

1. $S = \{(x,y,z) : x + 2y - \cancel{z} = 4\} \subset R^3$
2. $S = \{(x,y,z) : x + 2y - z \leq 4\} \subset R^3$
3. $S = \{(x,y) ; y \geq |x|\} \subset R^2$
4. $S = \{(x,y) : x^2 + y^2 \leq 4\} \subset R^2$

Lemma:- If S and T are convex sets, then

(i) $S \cap T$ is convex

(ii) $S + T$ is convex.

(iii) $S - T$ is convex.

Convex Combination:- Let $x_1, x_2, \ldots, x_n$ be in R^n. Then $x = \sum_{i=1}^{K} \lambda_i x_i$ is called a convex combination of the K points provided $\lambda_i \geq 0$ for all i and $\sum_{i=1}^{K} \lambda_i = 1$. If the non-negativity condition of λ_i's is dropped then the combination is called affine combination.

Convex function :- Let S be a nonempty convex set in R^n. A function $f: S \to R$ is said to be convex on S if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall\, x, y \in S,\ \forall\, \lambda \in (0,1)$$

Examples :- 1. $f(x) = 3x+4$ 2. $f(x) = |x|$

3. $f(x) = x^2 - 2x$ 4. $f(x) = x'Ax$, where A is psd.

The function $f(x) = -\sqrt{x}$ is a convex function on R_+.

Directional Derivative of Convex functions :-

Definition :- Let S be a nonempty set in R^n and let $f: S \to R$ be a function. Let $\bar{x} \in S$. A nonzero vector $d \in R^n$ is said to be a feasible ~~solution~~ direction of S at $\bar{x}$ if $\exists\ \delta > 0$ $\ni$ $\bar{x} + \lambda d \in S\ \forall\, \lambda \in (0, \delta)$.

Furthermore, for a feasible direction d of S at $\bar{x}$, f is said to have a directional derivative at $\bar{x}$ in the direction d if the following limit exists:

$$f(\bar{x}; d) = \lim_{\lambda \to 0^+} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda}$$

$f(\bar{x}; d)$ is the directional derivative of f at $\bar{x}$ in the direction d.

Lemma :- Let $f: R^n \to R$ be a convex function. Consider any point $\bar{x} \in R^n$ and a direction d. Then $f(\bar{x}; d)$ exists.

Proof :- Let $\lambda_2 > \lambda_1 > 0$,

$$f(\bar{x} + \lambda_1 d) = f\left[\frac{\lambda_1}{\lambda_2}(\bar{x} + \lambda_1 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right)\bar{x}\right]$$

$$\leq \frac{\lambda_1}{\lambda_2} f(\bar{x} + \lambda_1 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(\bar{x}) \quad [\text{By convexity of } f]$$

$\lambda_1 < \lambda_2$

$$\frac{f(\bar{x} + \lambda_1 d) - f(\bar{x})}{\lambda_1} \leq \frac{f(\bar{x} + \lambda_2 d) - f(\bar{x})}{\lambda_2}$$

Let $g(\lambda) = \dfrac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda}$. Then g is a nondecreasing function of λ over R_+.

Also by convexity of f, for any $\lambda > 0$,

$$f(\bar{x}) = f\left[\frac{\lambda}{1-\lambda}(\bar{x} - d) + \frac{1}{1-\lambda}(\bar{x} + \lambda d)\right]$$

$$\leq \frac{\lambda}{1-\lambda} f(\bar{x} - d) + \frac{1}{1-\lambda} f(\bar{x} + \lambda d)$$

Rearranging the terms, we get $g(\lambda) \geq f(\bar{x}) - f(\bar{x} - d)$.

Thus $g(\lambda)$ is bounded below, $\lim_{\lambda \to 0^+} g(\lambda)$ exists.

Subgradients of Convex functions:-

Let S be a nonempty convex set in R^n and let $f: S \to R$ be convex. Then a vector $\xi \in R^n$ is called a subgradient of f at a point $\bar{x} \in S$ if

$$f(x) \geq f(\bar{x}) + \xi^t (x - \bar{x}) \quad \forall \ x \in S.$$

Concave function & its subgradient :-

Let S be a nonempty convex set in R^n and let $f: S \to R$. Say that f is concave on S if $-f$ is convex. If f is a concave function, then a vector $\xi \in R^n$ is called a subgradient of f at a point $\bar{x} \in S$ if

$$f(x) < f(\bar{x}) + \xi^t (x - \bar{x}) \quad \forall \ x \in S.$$

Strictly Convex Function:- Let S be a nonempty convex set in R^n and let $f: S \to R$. Say that f is strictly convex on S if

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y) \quad \forall \ x, y \in S, \ x \neq y \text{ and } \forall \ \lambda \in (0,1)$$

Theorem:- Let S be a nonempty convex set of R^n and let $f: S \to R$. Suppose for every $x \in \text{int}(S)$, f has a subgradient ξ at x.

i.e. suppose for each $x \in \text{int}(S)$, $\exists$ a $\xi \ni f(u) \geq f(x) + \xi^t (u - x) \ \forall \ u \in S$.

Then f is convex on $\text{int}(S)$.

Proof:- Let $x, y \in \text{int}(S)$ and let $\lambda (0,1)$, $\bar{x} = \lambda x + (1-\lambda) y$ is in $\text{int}(S)$.

$$\exists \text{ a } \xi \ni \quad f(u) \geq f(\bar{x}) + \xi^t (u - \bar{x}) \ \forall \ u \in S. \quad \text{———} (*)$$

Note that

$$x - \bar{x} = (1-\lambda)(x - y)$$

$$y - \bar{x} = -\lambda (x - y).$$

Substituting x and y for u in $(*)$, we get

$$f(x) \geq f(\bar{x}) + (1-\lambda) \xi^t (x - y) \quad \text{———} \ ① \times (\lambda)$$

and

$$f(y) \geq f(\bar{x}) - \lambda \xi^t (x - y) \quad \text{———} \ ② \times (1-\lambda)$$

And then adding two will give

$$\lambda f(x) + (1-\lambda) f(y) \geq f(\bar{x}) = f(\lambda x + (1-\lambda) y).$$

Hence f is convex on $\text{int}(S)$.

Differentiable Convex Functions:-

Definition:- Let S be a set in R^n with nonempty interior and let $f: S \to R$. Let $\bar{x} \in \text{int}(S)$. Say that f is differentiable at $\bar{x}$ if $\exists$ a vector $\nabla f(\bar{x})$, called gradient vector of f at $\bar{x}$, and there exists a function $\alpha: R^n \to R$, $\ni$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x-\bar{x}) + \|x-\bar{x}\| \alpha(\bar{x}, x-\bar{x})$$

where $\lim_{x \to \bar{x}} \alpha(\bar{x}; x-\bar{x}) = 0$.

If T is a open subset of S, then f is said to be differentiable on T if f is differentiable at each point in T.

Remark:- $$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \frac{\partial f(\bar{x})}{\partial x_2}, \ldots\ldots, \frac{\partial f(\bar{x})}{\partial x_n}\right)^t$$

Theorem:- Let S be a nonempty open convex set of R^n and let $f: S \to R$ be differentiable on S. Then f is convex iff for every $\bar{x} \in S$

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x-\bar{x}) \quad \forall\ x \in S.$$

(Necessary & sufficient condition for a differentiable function to be convex)

Theorem:- Let S be a nonempty open convex set of R^n and let $f: S \to R$ be differentiable on S. Then f is convex iff for every $x, y \in S$, we have

$$\left| \nabla f(y) - \nabla f(x) \right|^t (y-x) \geq 0$$

[Characterisation of Convex function]

[For strict convex there is no equality in the expression]

Twice differentiable function:- Let S be a set in R^n with nonempty interior and let $f: S \to R$. Let $\bar{x} \in \text{int}(S)$. Say that f is twice differentiable at $\bar{x}$ if $\exists$ a vector $\nabla f(\bar{x})$, a symmetric mtx $H(\bar{x})$, called Hessian matrix and a function $\alpha: R^n \to R$, $\ni$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x-\bar{x}) + \frac{1}{2}(x-\bar{x})^t H(\bar{x})(x-\bar{x}) + \|x-\bar{x}\|^2 \alpha(\bar{x}, x-\bar{x});$$

for each $x \in S$ and $\lim_{x \to \bar{x}} \alpha(\bar{x}, x-\bar{x}) = 0$.

Remark:- When f is twice differentiable, $H(x)$ is given by

$$H(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x_1)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_n} \\ \cdots & \cdots & & \cdots \\ \frac{\partial^2 f(\bar{x}_0)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

Definition:- A square matrix is said to be positive semidefinite (positive definite) if $x^t A x \geq 0 \;\forall x$ ($x' A x > 0 \;\forall x \neq 0$)

Theorem:- Let S be a nonempty open convex set of R^n and let $f: S \to R$ be twice differentiable on S. Then f is convex iff the Hessian matrix of f is positive semidefinite at each point in S.

Note:- If $H(\bar{x})$ is positive definite, the f is strictly convex.

Maxima and minima of Convex Functions:- * Problem is: Minimize $f(x)$ sub $x \in S$

An $\bar{x} \in S$ is called global optimal solution to the to the problem if $f(x) \geq f(\bar{x}) \;\forall\; x \in S$.

An $\bar{x} \in S$ is called local optimal solution (minimum) if $\exists$ an $\epsilon > 0 \ni f(x) \geq f(\bar{x}) \;\forall\; x \in S$ with $\|x - \bar{x}\| < \epsilon$.

* Here $f: R^n \to R$ be a function, where $S \subseteq R^n$.

Prob. If f and S are convex, then show local optimal solution is global optimal solution.

Sol. $f: S \to R$ be convex function.
S: nonempty convex subset of R^n.
$\bar{x} \in S$ is a local optimal solution to the opt. prob.

We are to show $\bar{x}$ is also a global optimal solution.

To the contrary, assume that $f(y) < f(\bar{x})$ for some $y \in S$. For $\lambda \in (0,1)$, $\lambda y + (1-\lambda)\bar{x} \in S$, and by convexity of f,

$$f(\lambda y + (1-\lambda)\bar{x}) \leq \lambda f(y) + (1-\lambda) f(\bar{x})$$
$$< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x})$$

For λ sufficiently close to 0, $\lambda y + (1-\lambda)\bar{x} \in S$ can be made arbitrarily close to $\bar{x}$ which will contradict local optimality of $\bar{x}$.

Theorem:- (Necessary & Sufficient condition)
Let $f: R^n \to R$ be a convex function, where S is a nonempty convex subset of R^n. An $\bar{x} \in S$ is an optimal solution to the problem minimize $f(x)$ subject $x \in S$ iff, f has a subgradient ξ at $\bar{x}$ $\ni$ $\xi^t(x-\bar{x}) \geq 0$ $\forall x \in S$.

Cor:-
- If S is open nonempty convex subset of R^n, then $\xi = 0$.
- If f is differentiable function then $\nabla f(\bar{x})^t (x-\bar{x}) \geq 0$ $\forall x \in S$ will be the necessary & sufficient condition for $\bar{x}$ being optimal solution.

Theorem:- Consider the problem of minimizing $f(x)$ subject to $x \in S$ where f is convex function and two differentiable function, and S is a convex set, and suppose that $\exists$ an optimal solution $\bar{x}$, then, the set of alternative optimal solution to the problem will be

$$\gamma = \{ x \in S : \nabla f(x)^t (x-\bar{x}) \leq 0, \ \nabla f(x) = \nabla f(\bar{x}) \}$$

Quasiconvex Functions:- Let S be nonempty convex subset of R^n. A function from $f: S \to R$ is said to be quasiconvex if

$$f(\lambda x + (1-\lambda) y) \leq \max(f(x), f(y))$$
$$\forall x, y \in S, \forall \lambda \in [0,1]$$

The function f is said to be quasiconcave if $-f$ is quasiconvex.

Strict quasiconvex means $f(\lambda x + (1-\lambda)y) < \max(f(x), f(y))$

Theorem:- Let S be nonempty open convex subset of R^n. Let $f: S \to R$ be a function differentiable on S. Then f is quasiconvex iff the following holds good;

$$\forall x, y \in S, \ f(x) \leq f(y) \Rightarrow \nabla f(y)^t (x-y) \leq 0.$$

Strictly quasi-convex means $\longrightarrow$ f is continuous and quasi convex.

<u>Prob.</u>:- If f is strictly quasiconvex function and S is convex set, then s.t. local optimal solution is global optimal solution to the optimization problem.

<u>Proof</u>:- Let $\bar{x}$ be a local optimal solution to the problem. Suppose $y \in S$ is $\ni$ $f(y) < f(\bar{x})$, for any $\lambda \in (0,1)$ by strict convexity of f,

$$f(\bar{x} + \lambda(y - \bar{x})) < f(\bar{x}).$$

for λ sufficiently small, this will imply that $\bar{x} + \lambda(y - \bar{x})$ is locally better than $\bar{x}$, contradicting local optimality of $\bar{x}$.

<u>NOTE</u>:- f is strictly convex $\Rightarrow$ f is convex

f is strictly quasiconvex $\not\Rightarrow$ f is quasiconvex.

<u>Eg</u>:- Let $S = [-1, 1]$ and $f(x) = 0$, $\forall\, x \neq 0$ and $f(0) = 1$.

<u>Pseudoconvex function</u>:- Let S be a nonempty open convex set in $\mathbb{R}^n$ and let $f: S \to \mathbb{R}$ be a differentiable on S. Then the function f is said to be pseudoconvex on S if for each $x, y \in S$, the implication holds:

$$f(x) < f(y) \Rightarrow \nabla f(y)^t (x - y) < 0.$$

- f is differentiable pseduconvex $\Rightarrow$ f is quasiconvex + f is strictly quasiconvex.

<u>Ex.</u> $f(x) = (x^2 - 1)^3$; $\nabla f(x) = 6x(x^2 - 1)^2$; $\nabla f(y)^t (x - y) \geq 0 \Rightarrow f(x) \geq f(y)$

$\therefore$ f is pseudoconvex at $y = 0$.

Strictly pseduconvex means:

$$f(x) \leq f(y) \Rightarrow \nabla f(y)^t (x - y) < 0.$$

<u>Note</u>: 1. Every convex function is strictly quasiconvex as well as quasiconvex.

2. Every differentiable convex function is pseudoconvex.

TRANSPORTATION PROBLEM

Assumption:- Company has m depots (warehouses) and n destinations (Dealers). At each depot (i), the quantity available is a_i. At each destination (say, j) demand for the product is b_j. C_{ij} is the cost of transporting unit quantity from i to source destination j. Find the optimum allocation, such that overall cost of transportation is minimum and at the same time all demands are met.

Method of IBFS:-

(1) North West Method:- Find the N-W cell. Assign $\min\{a_i, b_j\}$. Cancel the row or column for which the constraints are satisfied. Find the N-W cell in ~~which~~ remainder matrix, follow this procedures till all assignments are done. Usually we get $n+m-1$ allocations.

Transportation Problem:-

x_{ij} : quantity transported from i to j

$$\text{Minimize } \sum\sum c_{ij}x_{ij}$$

$\sum a_i = \sum b_j$

$c_{ij} \geq 0$

$$\sum x_{ij} \leq a_i \ \forall i$$

$$\sum x_{ij} \geq b_j \ \forall j$$

$$x_{ij} \geq 0$$

balanced problem has equations, unbalanced problem has inequalities

- Identifying a B.F.S.
  - North West Corner Rule
  - Minimum Cost Method
  - Vogel's Approximation Method (VAM)

- Optimal Solution — MODI or U-V Method (Modified Distribution Method)

NW Method:-

Destination points

Supply Points					Supply
	4 20	6 20	8	8	~~40~~ 20
	6	8 10	6 50	7	~~60~~ 50
	5	7	6	8 50	50
Demand	20	~~30~~ 10	50	50	$\boxed{150}$

[5 allocations]

Cost of transportation = $4\times20 + 6\times20 + 8\times10 + 6\times50 + 8\times50$

$= 980.$

Minimum Cost Method:-

4 20	6 20	8	8	~~40~~ ~~20~~
6	8 X	6 50	7 10	~~60~~ ~~10~~
5	7 10	6	8 40	~~50~~ ~~40~~
20	~~30~~ 10	50	~~50~~ ~~40~~	

[6 allocations]

choose the minimum cost ij from the cells and assign min{ai, bi} and strike off the row/column.

$Z = 4\times20 + 6\times20 + 6\times50 + 7\times10 + 7\times10 + 8\times40$

$= 960.$

$m+n-1 = 6$

VAM :—

				Supply	Penalty
4 ✓ l.c.c 20	6 20	8	8	20 ~~40~~	2 ✓(highest penalty)
6	8 X	6 X 10	7 50	~~60~~ 50	~~0~~ 1
5	7 10	6 40	8 X	~~50~~ 40	~~1~~ ~~2~~ 1 2
20	~~30~~ 10	~~50~~ 10	50		
Penalty 1	1	0	1		

If $m+n-1 \neq 6$ then unbalanced case. Add another row or

[6 allocations]

$$Z = 4\times20 + 6\times20 + 6\times10 + 7\times50 + 7\times10 + 6\times40 = 920.$$

Method:-

1) Find the Penalties for each row & column.
   Penalty : | Difference between two least cost cells |
2) Determine the row/column where penalty is the highest.
3) In that row or column find the least cost cell.
4) Make the allocation $\min\{a_i, b_j\}$
5) Revise the matrix
6) Repeat it until all allocations are done.

Basic Feasible solution to a transportation problem:- Conditions are:

(i) The row-column (supply-demand) constraints are satisfied.
(ii) The non-negativity constrains are satisfied.
(iii) The allocations are dependent & don't form a loop.
(iv) There are exactly $m+n-1$ allocations.

We also observe that the three methods will not give more than $m+n-1$ allocations, since they allocate the maximum possible everytime, they will not have solutions with loops.

Table:-1

MODI METHOD

$u_i + v_j = c_{ij}$ for b.v.

$c_{ij} - (u_i + v_j)$ for N.b.v.

	$v_1 = 4$	$v_2 = 6$	$v_3 = 6$	$v_4 = 7$	
$u_1 = 0$	4 — 20	6 — 20	8 — (2)	8 — (1)	40
$u_2 = 0$	8 — (2)	8 — (2)	6 — 50 $-\theta$	7 — 10 $+\theta$	60
$u_3 = 1$	5 — (0)	7 — 10	6 — θ (−1)	8 — 40 $-\theta$	50
	20	30	50	50	

$\boxed{Z = 960}$

- Optimum solution is not reached.

$\min\{40, 50\} = 40$ add it to θ cells & substract from $-\theta$ cells

MODI Method:-

1. Introduce dual variables u_i's for a_i's and v_j's for b_j's $\ni$ $u_i + v_j = c_{ij}$ for the basic cells (where allocations are done).
2. Remember for minimisation problem, optimum solution is obtained when $z_j - c_j \le 0$
3. Verify if $z_j - c_j \le 0$ for all non-basic cells, i.e. $u_i + v_j - c_{ij} \le 0$.

Table:-2

MODI METHOD

	$v_1 = 4$	$v_2 = 6$	$v_3 = 5$	$v_4 = 6$	
$u_1 = 0$	$\boxed{4}$ 20	$\boxed{6}$ 20	$\boxed{8}$ (3)	$\boxed{8}$ (2)	40
$u_2 = 1$	$\boxed{6}$ (1)	$\boxed{8}$ (1)	$\boxed{6}$ 10	$\boxed{7}$ 50	60
$u_3 = 1$	$\boxed{5}$ (0)	$\boxed{7}$ 10	$\boxed{6}$ 40	$\boxed{8}$ (1)	50
	20	30	50	50	

$\boxed{Z = 920}$

Here the optimum solution is reached. $z_j - c_j \ge 0$.

4. Assign min{− sign cells} to + sign cells and subtract from + sign cells.
5. In addition to basic cells, if in case of non-basic cells $z_j - c_j = 0$, then the optimum solution is said to be cyclic in nature.

Scanned by CamScanner

ASSIGNMENT PROBLEM

→ Special case of transportation problem. Eg:- Job allocation

→ Have no. of jobs to be assigned to a no. of employees.

$$\text{Min} \sum\sum c_{ij} x_{ij}$$

$$\sum x_{ij} = 1 \ \forall j$$

$$\sum_j x_{ij} = 1 \ \forall i$$

$$x_{ij} = 0 \text{ or } 1$$

$$\begin{array}{c|cccc|} & 1 & 2 & 3 & 4 \\ \hline J1 & 5 & 9 & 3 & 6 \\ J2 & 8 & 7 & 8 & 2 \\ J3 & 6 & 10 & 12 & 7 \\ J4 & 3 & 10 & 8 & 6 \\ \hline \end{array}$$

$x_{ij} = 1$ if job i goes to person j

$= 0$ OW

$c_{ij} \geq 0$ if I have a feasible solution with $Z = 0$ unit is optimal solution.

To solve Assignment problem, we use Hungarian Method:-

1. <u>Row Reduction Method</u>:- Reduce the Given matrix by subtracting min elements of each row from all elements of that row.

2. <u>Column Reduction Method</u>:- Subtract the smallest element of each column from the elements of that column.

Row Reduction

$$\begin{bmatrix} 2 & 6 & 0 & 3 \\ 6 & 5 & 6 & 0 \\ 0 & 4 & 6 & 1 \\ 0 & 7 & 5 & 3 \end{bmatrix}$$

Column Reduction

$$\begin{bmatrix} 2 & 2 & \boxed{0} & 3 \\ 6 & 1 & 6 & \boxed{0} \\ 0 & \boxed{0} & 6 & 1 \\ \boxed{0} & 3 & 5 & 3 \end{bmatrix}$$

Optimum value of assignment $= 3 + 2 + 10 + 3$

$= 18$.

Ex.1

11	7	10	17	10
13	21	7	11	13
13	13	15	13	14
18	10	13	16	14
12	8	16	19	10

Row Reduction:-

4	0	3	10	3
6	14	0	4	6
0	0	2	0	1
8	0	3	6	4
4	0	8	11	2

Column Reduction:-

Cross max '0' lines

4	$\boxed{0}$	3	10	2	✓
6	14	$\boxed{0}$	4	5	
$\boxed{0}$	$\cancel{0}$	2	$\cancel{0}$	$\cancel{0}$	
8	$\cancel{0}$	3	6	3	✓
4	$\cancel{0}$	8	11	1	✓

$\theta = 1$

1. Tick the unassigned row
2. Tick the corresponding column if there is a zero.

~~3. If there~~

3. Look at all the columns if there is a tick mark, if there is an assignment in the column, then put a tick mark in the corresponding row.
4. Repeat above steps till no further marking can be done.
5. Draw lines through all unmarked rows and all marked column.
6. Find the smallest no. in the non underlined entries
7. Add θ if two lines pass through, subtract θ if no lines pass through, others remain unchanged.

3	$\boxed{0}$	2	9	1
6	15	$\boxed{0}$	4	5
$\boxed{0}$	1	2	$\cancel{0}$	$\cancel{0}$
7	$\cancel{0}$	2	5	2
3	$\cancel{0}$	7	10	$\boxed{0}$

Don't assign if there is an assignment in the corresponding row & column.

$\theta = 1$

2	0	1	8	0
6	16	$\boxed{0}$	4	5
$\boxed{0}$	2	2	0	0
6	$\boxed{0}$	1	4	1
3	1	7	10	$\boxed{0}$

$\theta = 1$

1	$\cancel{0}$	$\cancel{0}$	7	$\cancel{0}$
6	17	$\boxed{0}$	4	6
$\boxed{0}$	3	2	$\cancel{0}$	1
5	$\boxed{0}$	$\cancel{0}$	3	1
2	1	6	9	$\boxed{0}$

$\theta = 1$

$\boxed{0}$	$\cancel{0}$	$\cancel{0}$	6	$\cancel{0}$
5	17	$\boxed{0}$	3	6
$\cancel{0}$	4	3	$\boxed{0}$	2
4	$\boxed{0}$	$\cancel{0}$	2	1
1	1	6	8	$\boxed{0}$

Optimum value of assignment

$= 11 + 7 + 13 + 10 + 10$

$= 51$

Ex.

16	22	24	20
10	32	26	16
10	20	46	30

Soln:- Since the problem is of unbalanced assignment, so introducing new row with cost coefficient '0'.

$$\begin{bmatrix} 16 & 22 & 24 & 20 \\ 10 & 32 & 26 & 16 \\ 10 & 20 & 46 & 30 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Row Reduction:-

$$\begin{bmatrix} 0 & 6 & 8 & 4 \\ 0 & 22 & 16 & 6 \\ 0 & 10 & 36 & 20 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Column reduction will give the same result too.

Using Hungarian method:-

min. of unallocated place: $\theta = 4$

So, subtract it from unallocated and add to intersection.

$$\begin{bmatrix} 0 & 2 & 4 & \boxed{0} \\ \boxed{0} & 18 & 12 & 2 \\ 0 & 6 & 32 & 16 \\ 4 & \boxed{0} & 0 & 0 \end{bmatrix}$$

✓ since there is an assignment in ticked column

✓ unassigned row

zero in unassigned row

Draw lines through unmarked ~~unassigned~~ row and marked column.

optimum:-

$22 + 16 + 10$

$= 48$.

$$\begin{bmatrix} 2 & 2 & 4 & \boxed{0} \\ \boxed{0} & 16 & 10 & 0 \\ 0 & 4 & 30 & 14 \\ 6 & \boxed{0} & 0 & 0 \end{bmatrix}$$

$\theta = 2$

$$\begin{bmatrix} 2 & \boxed{0} & 2 & 0 \\ 0 & 14 & 8 & \boxed{0} \\ \boxed{0} & 2 & 28 & 14 \\ 8 & 0 & \boxed{0} & 2 \end{bmatrix}$$

Scanned by CamScanner

Travelling salesman Problem

'—' = '∞'

$$d_{ij} \quad \begin{array}{c|ccccc|} & 1 & & & & 5 \\ \hline 1 & - & & & & \\ & & - & & & \\ & & & - & & \\ & & & & - & \\ 5 & & & & & - \\ \hline \end{array}$$

$x_{ij} = 1$ if the salesman visits j immediately after i

$d_{ij} + d_{jk} > d_{ik}$
(Triangular Inequality)

$$\text{Min } \sum \sum d_{ij} x_{ij}$$

$$\sum_j x_{ij} = 1 \ \forall i$$

$$\sum_i x_{ij} = 1 \ \forall j$$

$$x_{ij} = 0, 1$$

Real life application of TSP:-

I. Machine Sequencing problems.
- Job scheduling
- Assembly line sequencing
- Cellular manufacturing

II. Sequencing Problem in Electronic Industry
- Drilling holes in Printed IC Board
- Testing IC via 'scan chain' technology.

III. Vehicle Routing for Delivery and despatch
- School Bus routing
- Parcel/Postal delivery
- Meals on wheel
- Maintainance on wheel.

IV. Genome sequencing for genetic study.

Integer Programming:-

LP Prob

+ x_j are integers.

1. – Linear – Non-linear
2. – All integer – zero-one

Let n be the no. of decision variable.
Suppose the first q $(1 \le q \le n)$ variables are constrained to be integer.

$$\text{Max } c^T x$$

$$\text{s.t. } Ax \le b$$

$$x \ge 0.$$

$c \rightarrow n \times 1$

$b \rightarrow m \times 1$

$A \rightarrow m \times n$

i.e. $\text{Max } \sum c_j x_j$ (z)

s.t. $\sum a_{ij} x_j \le b_i$

$x_j \ge 0$

$x_1, x_2, \ldots, x_q$ are integers.

$x_1, x_2, \ldots, x_q, \ldots, x_n$

if $q = 0$ it is a L.P.

$1 \le q \le n$ it is MIP

$q = n$ it is PIP.

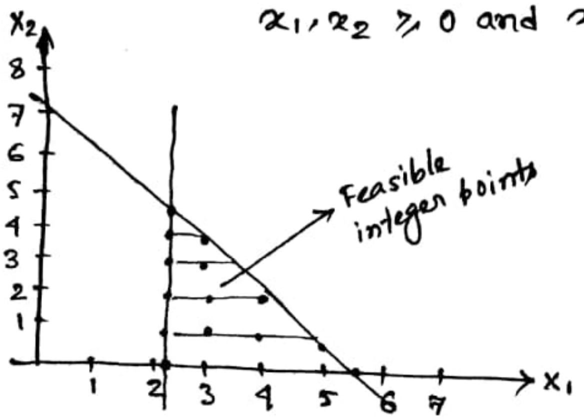
When $x_j \in \{0, 1\}$ it is binary programming.

Ex. Max $z = x_1 + x_2$

s.t. $15x_1 + 12x_2 \le 85$

$5x_1 \ge 11$

$x_1, x_2 \ge 0$ and x_1, x_2 are integers.



Optimum:-

$x_1 = 2\frac{1}{2}$, $x_2 = 4\frac{1}{3}$

$z =$

1. Branch and Bound Algorithm:-

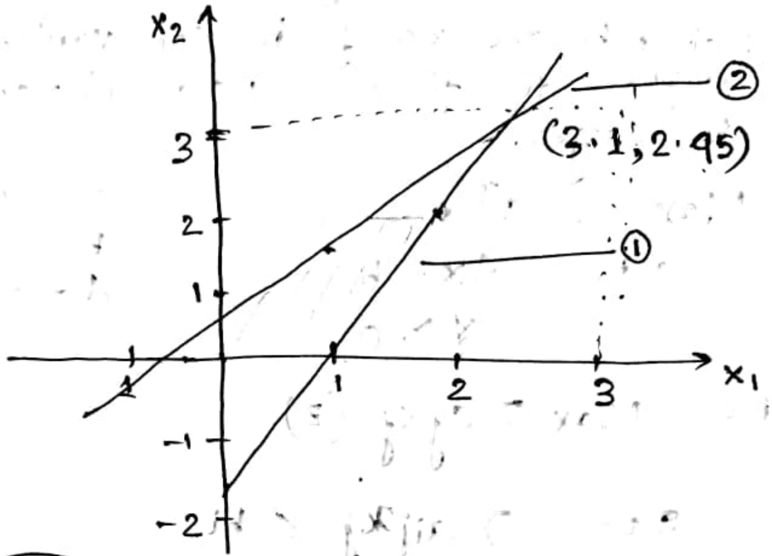
Ex.1. Maximize $x_1 + x_2$

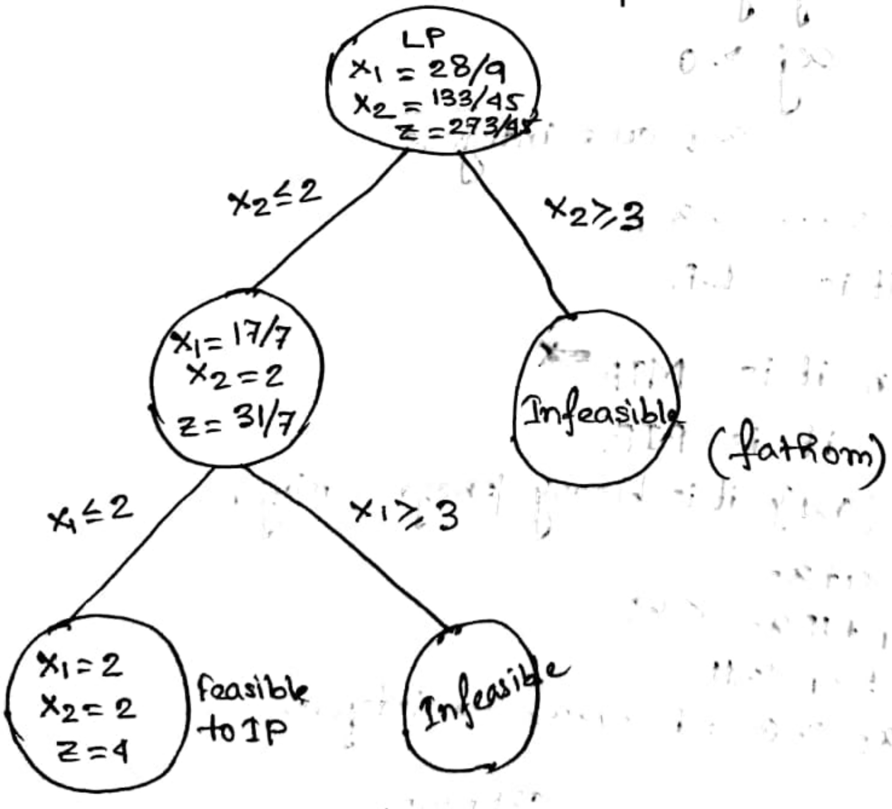
s.t. $7x_1 + 5x_2 \le 7$

$-12x_1 + 15x_2 \le 7$

$x_1, x_2 \ge 0$ & integers.

Sol.



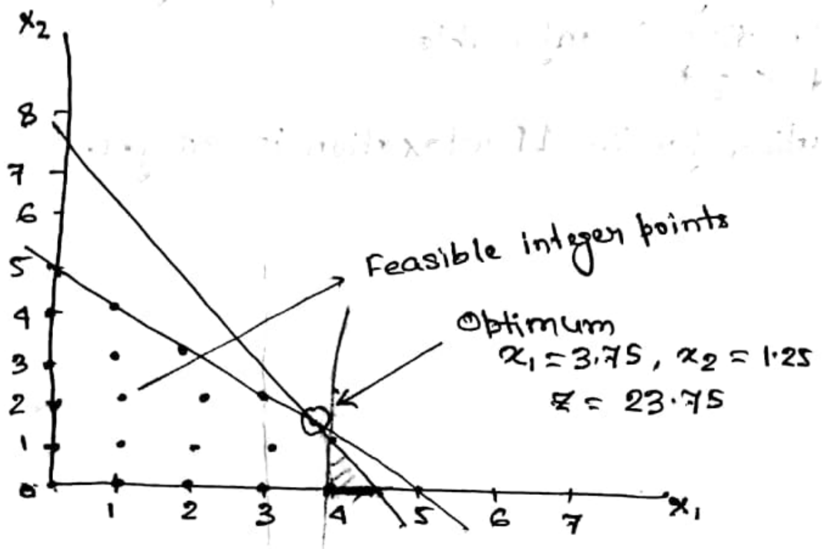


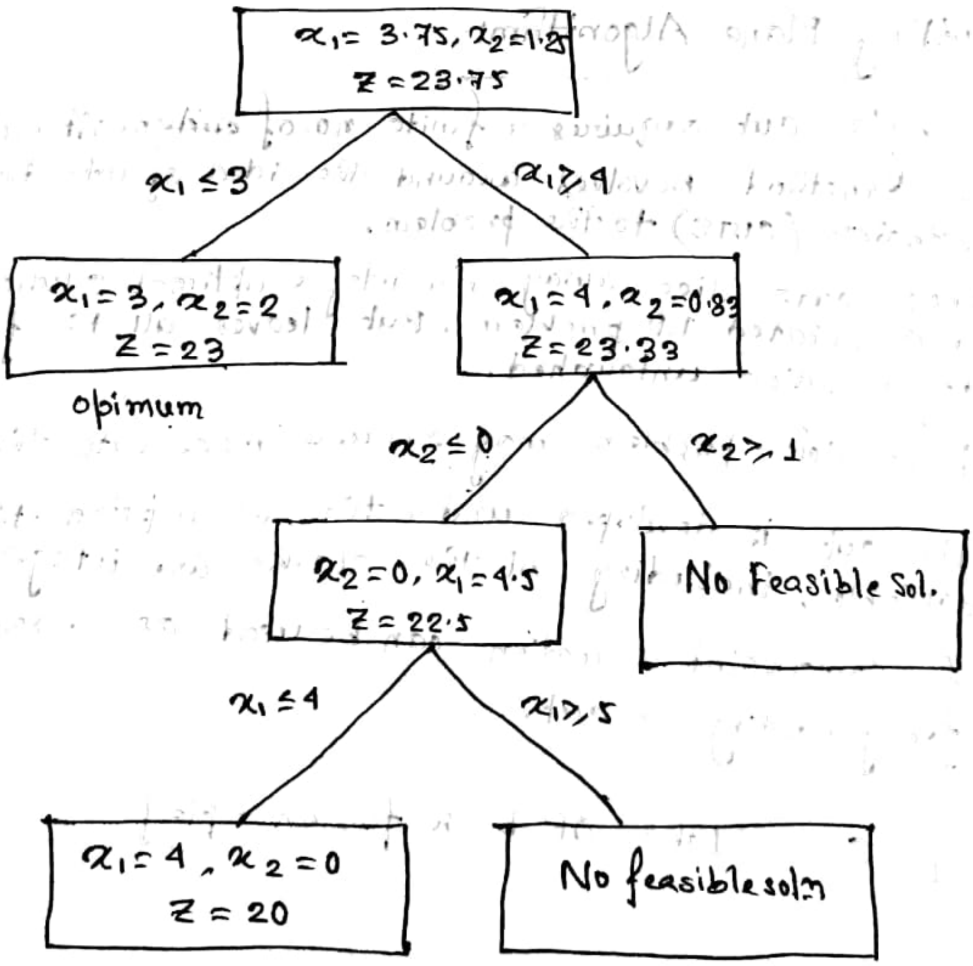
Ex.2. $\quad$ Max $Z = 5x_1 + 4x_2$

s.t. $x_1 + x_2 \le 5$

$10x_1 + 6x_2 \le 45$

x_1, x_2 are non-negative integer





- Branch and Bound is a sequential
  - → It is a kind of decision tree
  - → 1st iteration gives the root.

Note on fathoming criteria
- A subproblem is fathomed (dismissed from further consideration)

I. It is LP relaxation is infeasible
II. It is bound $\leq Z^*$
III. Optimal solution for its LP relaxation is integer.

2. Cutting Plane Algorithm:-

Gomory's cut requires a finite no. of iteration.

- The method revolves around the idea of introducing new constraints (CUTS) to the problem.
- These cuts slice away non-integer optimal solutions to the associated LP problem. But leaves all the feasible integer solution untouched.
- A smaller problem may require more cuts than larger one.
- The cut is developed under the assumption that all the variable, including all the slacks are integers.
- A constraint equation can be used as a source row for generating a cut.

[See Taha OR Book for example]

Capital Budgeting:- Decisions about whether or not to undertake a project made under-limited budget considerations & preset priorities.

Ex. (Project Selection) Five projects are being evaluated over a 3-year planning horizon. The following table gives the expected returns for each project & the associated yearly expenditures.

Project	Expenditure ($ million/year) 1	2	3	Returns ($ million)
1	5	1	8	20
2	4	7	10	40
3	3	9	2	20
4	7	4	1	15
5	8	6	10	30
Available Funds ($ million)	25	25	25	

Which projects should be selected over the 3-year horizon?

The problem reduces to a 'yes-no' decision for each project.

Define binary variable x_j as

$$x_j = \begin{cases} 1, & \text{if project } j \text{ is selected} \\ 0, & \text{if project } j \text{ is not selected} \end{cases}$$

The ILP model is

$$\text{Max } Z = 20x_1 + 40x_2 + 20x_3 + 15x_4 + 30x_5.$$

$$\text{s.t.} \quad 5x_1 + 4x_2 + 3x_3 + 7x_4 + 8x_5 \le 25$$

$$x_1 + 7x_2 + 9x_3 + 4x_4 + 6x_5 \le 25$$

$$8x_1 + 10x_2 + 2x_3 + x_4 + 10x_5 \le 25$$

$$x_1, x_2, x_3, x_4, x_5 = (0, 1).$$

The optimum integer solution is $x_1 = x_2 = x_3 = x_4 = 1, x_5 = 0$ (see next page).

The solution excludes project 5 from the product mix.

Job Sequencing Problem

Job	Processing time (day)	Due date (day)	Late penalty ($/day)
1	5	25	19
2	20	22	12
3	15	35	34

Define x_j = start date for job j

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ preceeds } j \\ 0 & \text{if } j \text{ preceeds } i \end{cases}$$

P_i and P_j processing time for job i & j

either $x_j \geq x_i + P_i$ or $x_i \leq x_j + P_j$

Either or constraint converted to

$$My_{ij} + (x_i - x_j) \geq P_j \quad (\text{M is very big no})$$

(since j^{th} job preceeds i^{th} job so $y_{ij} = 0$)

and $M(1-y_{ij}) + (x_j - x_i) \geq p_i$

Given that d_j is the due date for job j, the job is late if $x_j + p_j > d_j$. we can use two non-negative variables s_j^- and s_j^+ to determine the status of a completed job j with regards to its due date.

. Due date constraint can be written as

$$x_j + p_j + s_j^- - s_j^+ = d_j$$

Job j is ahead of schedule if $s_j^- > 0$.

Obj. functn. : $\text{Min } Z = 19\, s_1^+ + 12 s_2^+ + 34 s_3^+$

s.t.

$$x_1 - x_2 + My_{12} \geq 20$$
$$-x_1 + x_2 - My_{12} \geq 5 - M$$
$$x_1 - x_3 + My_{13} \geq 15$$
$$-x_1 + x_3 - My_{13} \geq 5 - M$$
$$x_2 - x_3 + My_{23} \geq 15$$
$$-x_2 + x_3 - My_{23} \geq 20 - M$$
$$x_1 + s_1^- - s_1^+ = 25 - 5$$
$$x_2 + s_2^- - s_2^+ = 20 - 22$$
$$x_3 + s_3^- - s_3^+ = 35 - 15$$

$x_1, x_2, x_3, s_1^-, s_1^+, s_2^+, s_2^-, s_3^+, s_3^- \geq 0$; $y_{12}, y_{23}, y_{13} = (0,1)$

This is a mixed ILP.

Application of Integer Programming:-

1. The knap Sack Problem:-

n : No. of different type of possessions to be taken
p_i : Value assigned to an integer of type i
f_i : The weight of an item of type i
c : Total weight that can be carried.
x_i : No. of items of type i to be taken.

So, objective function is $\sum x_i p_i$

s.t. $\sum f_i x_i \le c$.

2. The assignment Problem:-

n machine
n worker

c_{ij} : time taken by worker i to do the job in machine j.

Let $x_{ij} = \begin{cases} 1 & \text{if } i^{th} \text{ worker gets M/c } j \\ 0 & \text{Otherwise} \end{cases}$

Minimize $Z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij}$

s.t. $\sum_j x_{ij} = 1$ and $\sum_i x_{ij} = 1$.

* See 3 (TSP) in Page:- 25.

Note on Converting Mixed IP problem to Pure Binary IP:-

$x_1 \le 5$

$2x_1 + 3x_2 \le 30$

$$\begin{array}{r|l} 2 & 5 \\ \hline 2 & 2+1 \\ \hline & 1+0 \end{array}$$

$5 = 1\times 2^0 + 0\times 2^1 + 1\times 2^2$

$x_1 = y_1 + 2y_2 + 4y_3$

y_i are binary numbers.

In general, $x = 2^0 y_1 + 2^1 y_2 + \cdots + \sum_{i=1}^{n} 2^{i-1} y_i$

So,

$2(y_1 + 2y_2 + 4y_3) + 3(y_4 + 2y_5 + 4y_6 + 8y_7) \le 30$

$$\begin{array}{r|l} 2 & 10 \\ \hline 2 & 5+0 \\ \hline 2 & 2+1 \\ \hline & 1+0 \end{array}$$

$x_2 = y_4 + 2y_5 + 4y_6 + 8y_7$

4. Cutting Stock Problem:-

The determination of how to cut the larger standard sizes in to smaller 'ordered' sizes at minimum cost is called cutting stock problem.

Procedure:- Output parameters: standard weights $\omega_1, \omega_2, \ldots, \omega_m$

Decision variable: No. of standard rolls (y_j, T_j)

State variable: T_j (Trim loss); $j = 1(1)n$.

Constraint:- Total no. of each ordered width ω_i made must be at least b_i (customer demand).

Objective Function:- Minimize the no. of standard rolls needed

$$\text{Min} \sum_j c_j y_j$$

$$\text{s.t.} \sum_{j=1}^{n} a_{ij} y_j \geq b_i$$

$$y_j \geq 0 \text{ and integer.}$$

Problem:- Let us assume that there are K standard widths denoted by ω^K $\{\omega^1, \omega^2, \omega^3, \ldots, \omega^k\}$

$$\text{Min} \sum_{K=1}^{K} \sum_{j=1}^{n^K} c_j^K y_j^K$$

$$\text{s.t.} \sum_{K=1}^{K} \sum_{j=1}^{n^K} a_i^K y_j^K \geq b_i \quad , i = 1, 2, \ldots, m$$

$$y_j^K \geq 0 \text{ and integer.}$$

$j = 1, 2, \ldots, n^K$, $K = 1, 2, \ldots, k$; n is the number of cutting pattern

Dynamic Programming

Keywords:- Stage ; State ; Return

Return:- The return of a state is the minimum distance from hut 1 to the hut of that state.

Principal of optimality of Dynamic Programming:-

The decisions of the optimal policy for stages beyond a given stage will constitute an optimal subpolicy regardless of how the system entered that stage.

Basic Principle of counting:- When there are m ways to do one thing and n ways to do another, then there are m×n ways of doing both.

Forward Recursive Relation:-

d_{ij} : cost of transforming the system from state i to state j.

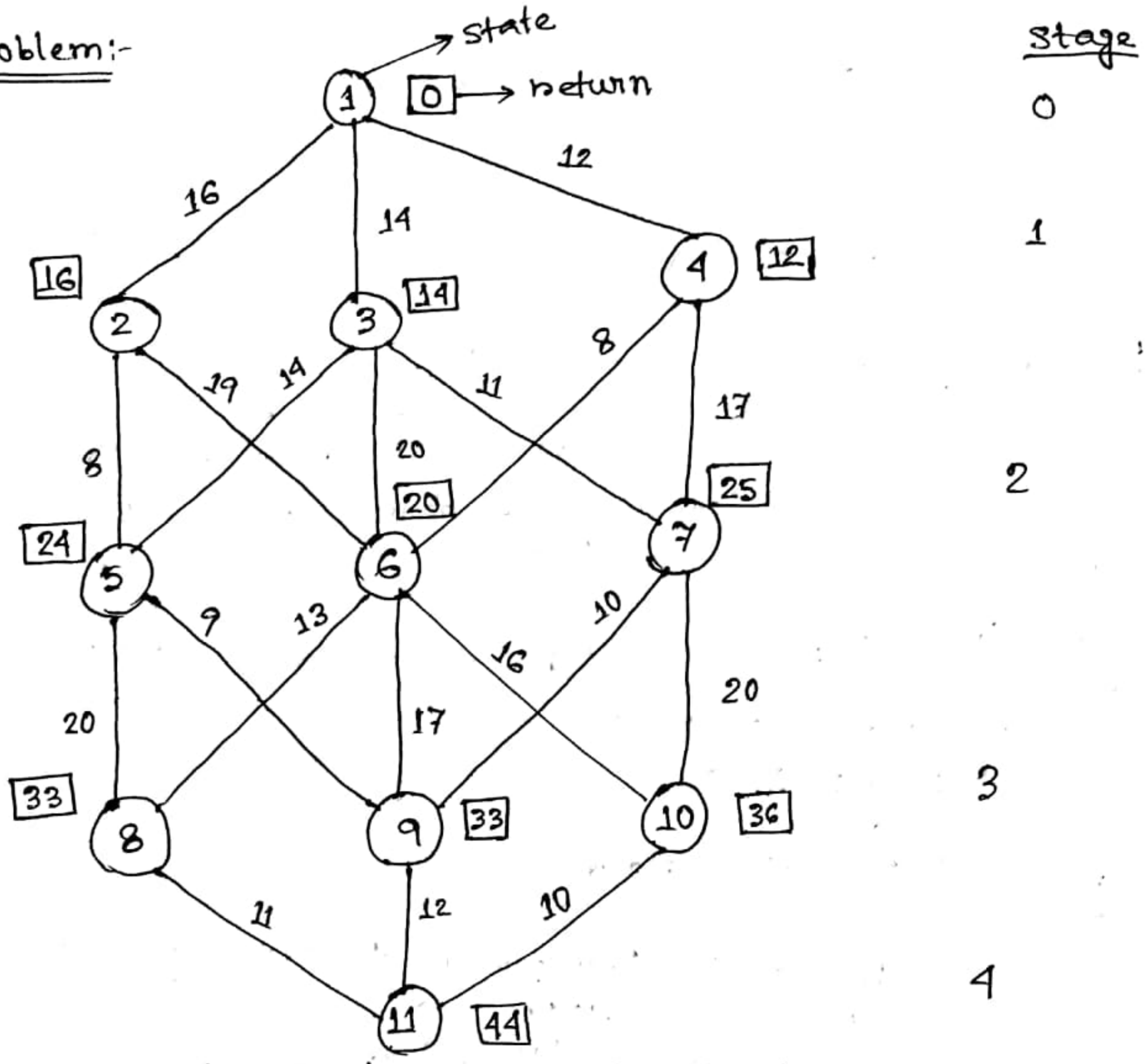
S_n : the state of the nth stage.

N = No. of stages of the system - 1.

$f_n(S)$ = Return of state S when the system is at stage n.

$$f_n(s) = \min \{ f_{n-1}(s_{n-1}) + d_{s_{n-1}, s} \}.$$

Problem:-



Forward Recursion:—

$f_0(s_1) = 0$

$f_1(s_2) = f_0(s_1) + d_{1,2} = 0 + 16 = 16$

$f_1(s_3) = f_0(s_1) + d_{1,3} = 0 + 14 = 14$

$f_1(s_4) = f_0(s_1) + d_{1,4} = 0 + 12 = 12$

$f_2(s_5) = \min\{ f_1(s_2) + d_{2,5},\ f_1(s_3) + d_{3,5} \}$

$= \min\{16 + 8,\ 14 + 14\} = 24$

$f_2(s_6) = \min\{ f_1(s_2) + d_{2,6},\ f_1(s_3) + d_{3,6},\ f_1(s_4) + d_{4,6} \}$

$= \min\{16 + 19,\ 14 + 20,\ 12 + 8\} = 20$

$f_2(s_7) = \min\{ f_1(s_3) + d_{3,7},\ f_1(s_4) + d_{4,7} \}$

$= \min\{14 + 11,\ 12 + 17\} = 25$

$$f_3(s_8) = \min\{f_2(s_5) + d_{5,8},\ f_2(s_6) + d_{6,8}\}$$

$$= \min\{24+20,\ 20+13\} = 33$$

$$f_3(s_9) = \min\{f_2(s_5) + d_{5,9},\ f_2(s_6) + d_{6,9},\ f_2(s_7) + d_{7,9}\}$$

$$= \min\{24+9,\ 20+17,\ 25+10\}$$

$$= 33.$$

$$f_3(s_{10}) = \min\{f_2(s_7) + d_{7,10},\ f_2(s_6) + d_{6,10}\}$$

$$= \min\{25+20,\ 20+16\}$$

$$= 36.$$

$$f_4(s_{11}) = \min\{f_3(s_8) + d_{8,11},\ f_3(s_9) + d_{9,11},\ f_3(s_{10}) + d_{10,11}\}$$

$$= \min\{33+11,\ 33+12,\ 36+10\}$$

$$= 44.$$

Optimum route:- 1 - 4 - 6 - 8 - 11

$12 + 8 + 13 + 11 = 44.$

Cargo-Loading Problem:- (Knapsack Model)

4 ton vessel, $W = 4$ ton

item i	w_i (weight)	r_i (revenue)
1	2	31
2	3	47
3	1	14

$f_i(m_i)$ = maximum return for stages $i, i+1, \ldots, n$

Let m_i be the no. of pieces of i^{th} item to be taken.

Objective function :

$$f_i(x_i) = \max_{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right]} \left\{ r_i m_i + f_{i+1}(x_{i+1}) \right\}, \quad i = 1, 2, \ldots, n$$

x_i : total weight assigned to stage $i, i+1, \ldots, n$.

s.t. $\sum m_i w_i \leq W$

$x_i - x_{i+1} = w_i m_i$; $x_{i+1} = x_i - w_i m_i$

So. $$f_i(x_i) = \max_{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right]} \left\{ r_i m_i + f_{i+1}(x_i - m_i w_i) \right\}$$

Stage-3 $\quad f_3(x_3) = \max_{m_3 = 0,1,2,4} \{14 m_3\}$

	$14m_3$					Optimum solution	
x_3	$m_3=0$	$m_3=1$	$m_3=2$	$m_3=3$	$m_3=4$	$f_3(x_3)$	m_3^*
0	0	–	–	–	–	0	0
1	0	14	–	–	–	14	1
2	0	14	28	–	–	28	2
3	0	14	28	42	–	42	3
4	0	14	28	42	56	56	4

Stage 2. $\max\{m_2\} = \left[\frac{4}{3}\right] = 1$ or $m_3 = 0, 1$

$$f_2(x_2) = \max_{m_3 = 0,1} \{47 m_2 + f_3(x_2 - 3m_2)\}$$

x_2	$47m_2 + f_3(x_2 - 3m_2)$		Optimum solution	
	$m_2 = 0$	$m_2 = 1$	$f_2(x_2)$	m_2^*
0	$0+0=0$	–	0	0
1	$0+14=14$	–	14	0
2	$0+28=28$	–	28	0
3	$0+42=42$	$47+0=47$	47	1
4	$0+56=56$	$47+14=61$	61	1

Stage 1. $\max\{m_1\} = \left[\frac{4}{2}\right] = 2$ or $m_1 = 0, 1, 2,$

$$f_1(x_1) = \max_{m_3 = 0,1,2} \{31 m_2 + f_2(x_1 - 2m_1)\}$$

x_1	$31m_1 + f_2(x_1 - 2m_1)$			Optimum solution	
	$m_1 = 0$	$m_1 = 1$	$m_1 = 2$	$f_1(x_1)$	m_1^*
0	$0+0=0$	–	–	0	0
1	$0+14=14$	–	–	14	0
2	$0+28=28$	$31+0=31$	–	31	1
3	$0+47=47$	$31+14=45$	–	47	0
4	$0+61=61$	$31+28=59$	$62+0=62$	62	2

$m_1 = 2$, $m_2 = m_3 = 0$

Max. Revenue $= 62$

PERT/CPM

PERT :- Project Evaluation & Review Techniaue

CPM :- Critical Path method

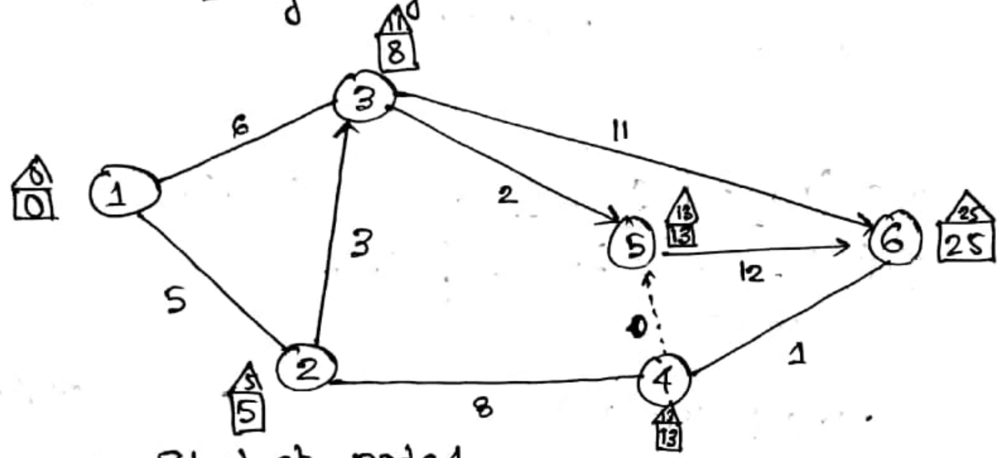
Planning, Scheduling, Executing, Control. Project

$\square_j \rightarrow$ earliest occurance time of event j

$\triangle_j \rightarrow$ latest occurance time of event i

$d_{ij} \rightarrow$ duration of activity (i,j)

$= \triangle_j - \square_j$



Start at node 1.

forward pass:-

$\square_1 = 0$

$\square_2 = 0 + 5 = 5$

$\square_3 = \max\{\square_1 + d_{13}, \square_2 + d_{23}\} = \max\{6, 5+3\} = 8$

$\square_4 = \square_2 + d_{24} = 5 + 8 = 13$

$\square_5 = \max\{\square_3 + d_{3,5}, \square_4, d_{4,5}\} = \max\{8+2, 13+0\}$
$= 13.$

$\square_6 = \max\{8+11, 13+12, 13+1\}$
$= 25$

Backward Pass:-

$\triangle_6 = \square_6$

$\triangle_5 = \triangle_6 - d_{5,6} = 25 - 12 = 13$

$\triangle_4 = \min\{\triangle_6 - d_{6,4}, \triangle_5 - d_{5,4}\}$
$= \min\{25 - 1, 13 - 0\} = 13$

$\triangle_3 = \min\{\triangle_6 - d_{6,3}, \triangle_5 - d_{5,3}\}$
$= \min\{25 - 11, 13 - 2\} = 11$

$$\Delta_2 = \min\{\Delta_3 - d_{2,3}, \Delta_4 - d_{2,4}\}$$

$$= \min\{11-3, 13-8\} = 5$$

$$\Delta_1 = \min\{\Delta_3 - d_{1,3}, \Delta_2 - d_{1,2}\}$$

$$= \min\{11-6, 5-5\} = 0$$

Critical path : 1-2-4-5-6.

Forward Pass formula:-

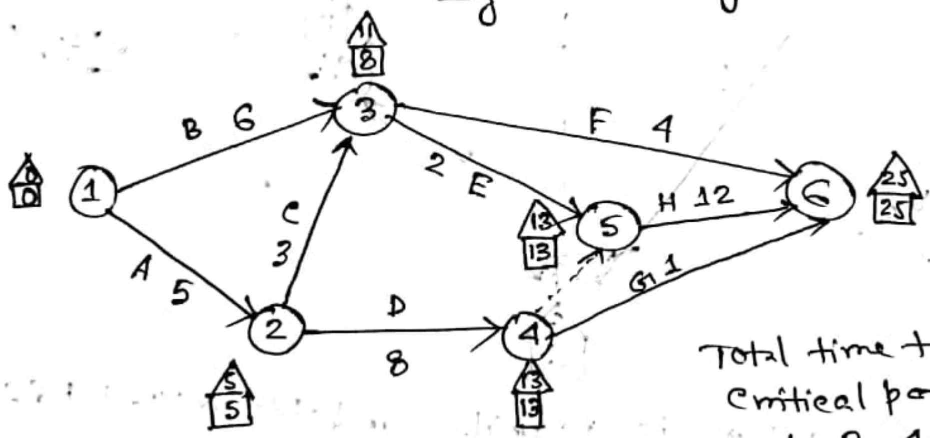
$$\square_j = \max\{\square_p + d_{pj}, \square_q + d_{q,j}, \ldots, \square_r + d_{rj}\}$$

Backward Pass formula:-

$$\Delta_j = \min\{\Delta_p - d_{jp}, \Delta_q - d_{jq}, \ldots, \Delta_r - d_{jr}\}$$

Free Float $= \square_j - \square_i - D_{ij}$

$\leq$ Total Float $= \Delta_j - \square_i - D_{ij}$



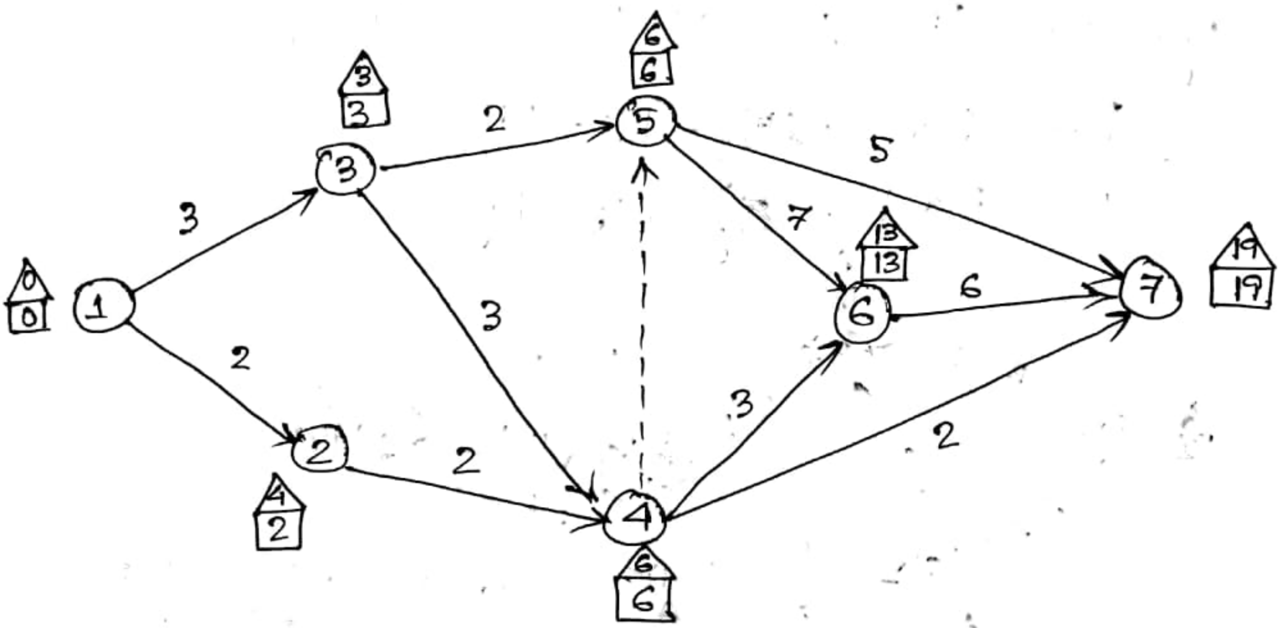
Total time taken is 25
Critical path is .
1-2-4-5-6

Activity	Duration	Total Float	Free Float
A	5	0	0
B	6	5 (11-0-6)	2 (8-0-6)
C	3	3	0
D	8	0	0
E	2	3	3
F	11	6	6
G	1	11	11
H	12	0	0

Red Flagging Rule:- For a NC activity (i,j) if $FF_{ij} < TF_{ij}$

Then its start can be delayed by at most FF_{ij}, relative to its earliest start time $\square_i$ without schedule conflict. Any delay larger than FF_{ij} must be coupled with an equal delay (relative to $\square_j$) in the start time of all the activities leaving node.

Ex.



Total time taken is 19 units and the critical path is

$1 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$.

[Total:- 100 Marks]

Final Exam:-

✓1. Integer Programming (Only Modelling) (Pg: 30-31) [20 Marks]

✓2. PERT / CPM

✓3. Queuing theory (1st four model) [20 Marks]

✓4. Inventory Theory (3 things) [20 Marks]

✓5. Simulation [20 Marks]

✓6. Multi-objective Linear Programming

PERT/CPM See Page:- 52-53

✓7. Goal Programming [20 marks]

QUEUEING THEORY

Queue → associate "measure of performance" of Queueing system.

Elements of a Queueing System:- Nature of arrival/departure/ service time are random.

Type of Queue:- – Parallel Queue (more than one counter)
- series Queue (checking in airport, one followed by another)
- Network Queue

Service Pattern:- – FIFO (First in First out)
- LIFO (Last in First out)
- Random
- Priority

Customer Behaviour:- – Jockey (Jumping from one queue to another short queue
- Balk (Leaving the queue after some time)
- Renege (Not come in the queue since queue size is long)

Q-size:- – Finite (Job interview)
- Infinite (Bank)

Measure of Performance:- – Expected Q-length
- Average waiting time
- % of time server in idle.

Probability Distr. of Inter-arrival time:-

An RV T is said to have Exponential distr. if

$$f(t) = \lambda e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}$$

$$E(t) = \frac{1}{\lambda} \quad ; \quad \lambda = 0.2 \text{ breakdown/hr}$$

then $f(t) = 0.2 e^{-0.2t}$

Loss of Memory Property: $P(T > t+s \mid T > s)$

$$= \frac{P(T > t+s)}{P(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}$$

$$= P(T > t)$$

$$= e^{-\lambda t}$$

Scanned by CamScanner

Ex. One electric bulb fails every month. What's the prob. that one electric bulb that was purchased on 28^{th} February, 2015 fail before 10^{th} April − 4pm? (Time now is 4pm - 30th March?

$$\lambda = \frac{1}{30}$$

Sol. $P(T > 10)$

$$= e^{-\frac{1}{30} \times 10} = 0.6930$$

$$P(T < 11) = 1 - P(T > 10) = 1 - 0.6930 = 0.307$$

- **Pure-Birth Model** (Random arrival to join the Queue)

Define $P_0(t)$ = Prob. of no arrival during a period of time t.
Given that inter arrival time is exponential and arrival rate is λ customers per unit time.

$$P_0(t) = P(\text{inter arrival time} \geq t)$$
$$= 1 - P(IAT < t)$$
$$= e^{-\lambda t}$$

Assume 'h' to be a very small time interval

$$P_0(h) = e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2!} \cdots\cdots = 1 - \lambda h + O(h^2)$$

$$\lim_{h \to 0} P_0(h) = 1 - \lambda h$$

$$P_1(h) = 1 - P_0(h) = 1 - (1 - \lambda h) = \lambda h .$$

Now, assuming that inter arrival time is exponential with mean $\frac{1}{\lambda}$.

Define $P_n(t)$ = Prob. of n arrivals during t.

$$P_n(t+h) \approx P_n(t)(1 - \lambda h) + P_{n-1}(t)\,\lambda h \quad \text{when } n > 0$$

$$P_0(t+h) \simeq P_0(t)(1 - \lambda h) \quad \text{when } n = 0$$

Rearranging the term, we get

$$P_n'(t) = \lim_{h \to 0} \frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t), \; n > 0$$

$$P_0'(t) = \lim_{h \to 0} \frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) \quad , n = 0 .$$

where $P_n(t) = \dfrac{\lambda t \, e^{-\lambda t}}{n!}$.

• **Pure Death Model:-** We start at time 0 with N person in the Queue. No arrival occurs and departure occurs at the rate of μ customers/unit time.

Assumption:- On a small time gap 'h' maximum 1 departure happens.

$P('0' \text{ departure}) = 1-\mu h$

$P('1' \text{ departure}) = \mu h$

Again our objective is to derive the prob. of n customers at time t, starting with N customers at time '0' and no arrival occurs.

$$P_N(t+h) = P_N(t)(1-\mu h)$$

$$P_n(t+h) = P_n(t)(1-\mu h) + P_{n+1}(t)\mu h$$

$$\lim_{h\to 0} \quad P_N'(t) = -\mu P_N(t)$$

$$P_n'(t) = -\mu P_n(t) + \mu P_{n+1}(t)$$

$$P_0'(t) = \mu P_1(t).$$

All these differential ~~difference~~ equation leads to the outcome

$$P_n(t) = \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)!} \quad ; \quad n = 1, 2, \dots, N$$

$$P_0(t) = 1 - \sum_{n=1}^{N} P_n(t).$$

Ex. In a city babies are born every 12 min

(i) Average No. birth per year $= 43800 = \lambda t = 120 \times 365$

$\lambda = 120$ birth/day $= \frac{24 \times 60}{12}$

(ii). P(No. birth in one day period) $= e^{-\lambda t} = e^{-120 \times 1} = 0.$

(iii) $\lambda =$ hourly birth rate $= \frac{60}{12} = 5$

P(10 births in one hour) $= \frac{(\lambda t)^{10} e^{-\lambda t}}{10!} = \frac{5^{10} e^{-5}}{10!}$

• **Ques (Pure-birth model):-** Babies are born in a large city at the rate of one birth every 12 minutes. The time between births follows Exp. distn. Find

(a) The average number of births per year.

(b) The prob. that no births will occur during 1 day.

(c) the prob. of issuing 50 birth certificates in 3hrs, given that 40 certificates were issued during the first 20hrs of the 3-hr period.

Ex. In a bank operation, the arrival rate is 2 customer/min;
Determine —
(a) The average no. of arrivals during a 5 minute;
(b) Prob. that no arrivals will occur during the next 0·5 minute
(c) Prob. that at least one arrival will occur during next 0·5 min.
(d) Prob. that the time between two successive arrivals is at least 3 minute.

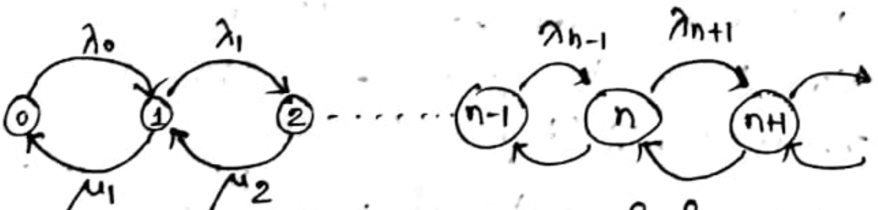
Ans. $\lambda = 2$ customer/min
(a) $P(\text{no of arrivals during 5min}) = 1 - e^{-\lambda t}$
(b) $P_1 = e^{-\lambda t}$
(c) $P = 1 - e^{-\lambda t} = 1 - P_1$
(d) $P(\text{no. interval in 3 min}) = e^{-\lambda t}$.

- **General Poisson Queuing:-**

1. It combines both pure birth & pure death model.
2. It assumes a steady state condition.
3. Arrival and departures are state dependent

Define, n = No. of customers in the system
λ_n = Arrival rate, given n customers in the system
μ_n = Departure rate, " " " " "
P_n = steady state probability

Transition diagram:-



In steady state expected rates of flow into and out of state n must be equal.
Expected rate of flow into state $n = \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}$
Expected rate of flow out state $n = (\lambda_n + \mu_n) P_n$.
Following balance equation due to steady state

$$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n, \quad n = 1, 2, \ldots$$

And for $n = 0$, $\lambda_0 P_0 = \mu_1 P_1$

By solving the recursive equations, we have

$$P_1 = \left(\frac{\lambda_0}{\mu_1}\right)P_0$$

and $\lambda_0 P_0 + \mu_2 P_2 = (\lambda_1 + \mu_1)P_1 = \frac{\lambda_1 + \mu_1}{\mu_1} \cdot \frac{\lambda_0}{\mu_1} \cdot P_0$

$$P_2 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} \cdot P_0$$

Continuing this method of mathematical induction

$$P_n = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_1} \cdot P_0$$

The value of P_0 is obtained from $\sum_{n=0}^{\infty} P_n = 1$.

Ex.

No. of customer in store	No. of cust
1 to 3	1
4 to 6	2
more than 6	3

Customer arrive according to a Poisson distribution with a mean rate of 10/hr. the average checkout time per customer is exponential with mean 12 minutes. Determine the steady state probability?

Ans. $\lambda_n = \lambda = 10$

$\mu_1 = 5;\ \mu_2 = 10,\ \mu_3 = 15$

$$P_1 = \frac{\lambda_1}{\mu_1} P_0 = 2P_0$$

$$P_2 = \frac{\lambda_2 \lambda_1}{\mu_2 \mu_1} P_0 = \frac{10 \times 10}{10 \times 5} P_0 = 2P_0$$

$$P_3 = \frac{\lambda_3 \lambda_2 \lambda_1}{\mu_3 \mu_2 \mu_1} = \frac{10 \times 10 \times 10}{15 \times 10 \times 5} P_0 = \frac{4}{3} P_0$$

$$P_0 + P_2 + P_3 = 1$$

$$\Rightarrow P_0 = \frac{3}{16}$$

General characteristic of a Queueing system:-

$$(a / b / c) / (d / e / f)$$

- a: arrival pattern
- b: service pattern
- c: no. of server
- d: service time
- e: system size
- f: size of the source

L_q : expected no. of person in the queue

L_s : expected no. of person in the system

W_q : Expected ~~no. of~~ waiting time in the queue

W_s : expected waiting time in the system

$\bar{c}$: busy server

λ_{eff} : effective arrival rate

$$L_s = \sum_{n=1}^{\infty} nP_n \quad ; \quad L_q = \sum_{n=c+1}^{\infty} (n-c)P_n$$

The relationship between L_s and W_s is

$$L_s = \lambda_{eff} \cdot W_s$$

$$L_q = \lambda_{eff} \cdot W_q$$

The relationship between W_s and W_q is

$$W_s = W_q + \frac{1}{\mu} \quad ; \quad \mu = \text{server rate}$$

$$L_s = L_q + \frac{\lambda_{eff}}{\mu},$$

Average (expected no. of busy server) $= L_s - L_q = \bar{c} = \frac{\lambda_{eff}}{\mu}$.

Facility Utilization $= \dfrac{\text{expected no. of busy servers}}{\text{total no. of servers}}$

Ex. Maximum Queue length $= 3$

No. of servers, $c = 5$

Max. system size $= 3+5 = 8$

Cars arrive according Poisson at 6cars/hr.

$\lambda = 6$

Parking time is exponential within meantime, $t = 30$min

$$\mu = \frac{60}{30} = 2 \text{cars/hr}.$$

If there are n parking place occupied, then $\mu = 2n$ car/hr.

$\lambda = 6\, car/hr$

$\mu_n = 2n$ for $n = 1, 2, \dots, 5$

$= 2 \times 5$ for $n = 6, 7, 8$

$P_n(t) = \frac{3^n}{n!} \cdot P_0$; $n = 1, 2, 3, 4, 5$.

1. **Single Server Model :-** $(M/M/1)/(GD/\infty/\infty)$

M : Markovian Poisson

GD : General Discipline

1 : Single Server

$\lambda_n = \lambda$ for $n = 0, 1, \dots$

$\mu_n = \mu$ for $n = 0, 1, \dots$

Also, $\lambda_{eff} = \lambda$ and $\lambda_{lost} = 0$.

$P_n = \frac{\lambda^n}{\mu^n} P_0 = \left(\frac{\lambda}{\mu}\right)^n P_0$; $\frac{\lambda}{\mu} = \frac{\text{arrival rate}}{\text{service rate}}$

$\therefore P_n = \rho^n P_0$

To determine P_0, we use $\sum_{n=0}^{\infty} P_n = 1$

$\sum_{n=0}^{\infty} \rho^n P_0 = 1 \Rightarrow P_0 = 1 - \rho$.

$\therefore P_n = \rho^n P_0 = (1-\rho)\rho^n$; $n = 1, 2, \dots$; $\rho < 1$

Expected length in the system,

$$L_s = E(P_n) = \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n$$

$$= (1-\rho)(\rho + 2\rho^2 + 3\rho^3 \dots)$$

$$= \rho(1-\rho)(1 + \rho^2 + 2\rho^3 + \dots)$$

$$= \rho(1-\rho)(1-\rho)^{-2}$$

$$= \frac{\rho}{1-\rho}.$$

$L_q = \lambda W_q$

$W_q = W_s - \frac{1}{\mu}$

$W_s = \frac{L_s}{\lambda}$.

Ex. Car wash: 1 server

$\lambda = 4$ cars/hour (arrival rate)

$\mu = 6$ cars/hour (service rate)

$$\rho = \frac{\lambda}{\mu} = \frac{4}{6} = \frac{2}{3}$$

$$L_s = \frac{\rho}{1-\rho} = \frac{2/3}{1-2/3} = 2.$$

$$W_s = \frac{L_s}{\lambda} = \frac{2}{4} = \frac{1}{2}$$

$$W_q = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$

$$L_q = \lambda W_q = 4 \times \frac{1}{3} = 1.33.$$

2. Model:- $(M/M/1)(GD/N/\infty)$

$$\lambda_n = \begin{cases} \lambda & ; n = 0, 1, \ldots, N-1. \\ 0 & ; n = N, N+1, \ldots \end{cases}$$

$$\mu_n = \mu \quad ; n = 0, 1, \ldots$$

$$P_n = \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1} P_0 = \frac{\lambda^n}{\mu^n} P_0 = \left(\frac{\lambda}{\mu}\right)^n P_0 = \rho^n P_0 \text{ for } n = 0, 1, \ldots, N$$

$$= 0 \qquad \text{for } n > N$$

$$P_n = \rho^n P_0$$

$$\sum_{n=0}^{N} P_n = 1 \Rightarrow P_0 \sum_{n=0}^{N} \rho^n = 1$$

$$\Rightarrow P_0 \left(1 + \rho + \rho^2 + \cdots + \rho^N\right) = 1$$

$$\Rightarrow P_0 \left[\frac{1-\rho^{N+1}}{1-\rho}\right] = 1 \text{ when } \rho \neq 1$$

when $\rho = 1$, $P_0 (N+1) = 1$.

$$P_n = \rho^n P_0 = \rho^n \left(\frac{1-\rho}{1-\rho^{N+1}}\right) \text{ when } \rho \neq 1$$

$$= \frac{1}{N+1} \quad ; \quad \rho = 1$$

$$\lambda_{lost} = \lambda P_N$$

$$\lambda_{eff} = \lambda(1 - P_N)$$

$$\lambda = \lambda_{lost} + \lambda_{eff}$$

The expected no. of customer in the system.

$$L_s = \sum_{n=1}^{N} n P_n$$

$$= \frac{1-\rho}{1-\rho^{N+1}} \sum_{n=1}^{N} n \rho^n$$

$$= \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho \sum_{n=0}^{N} \frac{d}{d\rho} [\rho^n]$$

$$= \frac{(1-\rho)\rho}{1-\rho^{N+1}} \frac{d}{d\rho} \left[\frac{1-\rho^{N+1}}{1-\rho} \right]$$

$$= \frac{\rho \left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)(1-\rho^{N+1})}, \rho \neq 1.$$

$$W_q = \frac{L_q}{\lambda_{eff}} = \frac{L_q}{\lambda(1-P_N)}$$

$$L_s = L_q + \frac{\lambda_{eff}}{\mu} = L_q + \frac{\lambda(1-P_N)}{\mu}$$

$$W_s = W_q + \frac{1}{\mu} = \frac{L_s}{\lambda(1-P_n)}$$

When $\rho = 1$, $L_s = \frac{N}{2}$.

Scanned by CamScanner

3) $(M/M/c)\ (GD/\infty/\infty)$, $c < N$

Starting with the general model,

$$P_n = \frac{\lambda_{n-1}\ \lambda_{n-2} \cdots \lambda_0}{\mu_n\ \mu_{n-1} \cdots \mu_1} \cdot P_0$$

$$\lambda_n = \begin{cases} \lambda & , 0 < n < N \\ 0 & , n > N \end{cases}$$

$$\mu_n = \begin{cases} n\mu & , 0 \le n < c \\ c\mu & , c \le n < N \end{cases}$$

$$\therefore P_n = \frac{\lambda\lambda \cdots \lambda}{\mu \cdot 2\mu \cdot 3\mu \cdots n\mu} P_0 = \frac{\lambda^n}{\mu^n \cdot n!} P_n \quad \text{when } n < c$$

When $c \le n < N$, $P_n = \dfrac{\lambda^n}{\mu^n c!\ c^{n-c}} P_0$

4) (Self Service Model) $(M/M/\infty)\ (GD/\infty/\infty)$

$$\lambda_n = \lambda, \quad n = 0, 1, \ldots, n.$$

$$\mu_n = n\mu,$$

Using General formula,

$$P_n = \frac{\lambda \cdot \lambda \cdots \lambda}{\mu \cdot 2\mu \cdots n\mu} P_0 = \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!} P_0 = \frac{\rho^n}{n!} P_0$$

$$\sum_{n=1}^{\infty} P_n = 1 \Rightarrow \sum \frac{\rho^n}{n!} P_0 = 1$$

$$\Rightarrow e^{\rho} P_0 = 1$$

$$\Rightarrow P_0 = e^{-\rho}.$$

$$\therefore P_n = \frac{e^{-\rho} \rho^n}{n!} \quad \text{(Poisson)}$$

PERT/CPM (Continued)

Activity	$i-j$	(a, m, b)	$\bar{D}_{ij} = \frac{a+b+4m}{6}$	$V_{ij} = \frac{(b-a)^2}{6^2}$
A	1-2	(3, 5, 7)	5	0.444
B	1-3	(4, 6, 8)	6	0.444
C	2-3	(1, 3, 5)	3	0.444
D	2-4	(5, 8, 11)	8	1
E	3-5	(1, 2, 3)	2	0.111
F	3-6	(9, 11, 13)	11	0.444
G	4-6	(1, 1, 1)	1	0
H	5-6	(10, 12, 14)	12	0.444

a = optimistic time
b = pesimistic time
m = most likely time

Node j	longest path	Path Mean	Path SD	S_j	K_j	$P(Z \le K_j)$
2	1-2	5	$\sqrt{0.444} = 0.67$	5	0	0.50
3	1-2-3	8	$\sqrt{0.444+0.444} = 0.94$	11	3.19	0.99
4	1-2-4	13	$\sqrt{0.444+1} = 1.20$	12	0.83	0.79
5	1-2-4-5	13	$\sqrt{0.444+1} = 1.20$	14	0.83	0.79
6	1-2-4-5-6	25	$\sqrt{0.444+1+0.444} = 1.37$	26	0.73	0.76

$$P(e_j \le S_j) = P\left(\frac{e_j - E(e_j)}{\sqrt{Var(e_j)}} \le \frac{S_j - E(e_j)}{\sqrt{Var(e_j)}}\right)$$

$$= P(Z \le K_j)$$

where, e_j = total expected time from node 1 to node j along the longest path.

Simulation

PCs demanded per week	Freq. of Demand	Prob. of Demand ($P(x)$)	Ranges of Random Nos
0	20	.20	0-19
1	40	.40	20-59
2	20	.20	60-79
3	10	.10	80-89
4	10	.10	90-99
Total	100	1	

Week	rn	demand, x	Revenue
1	39	1	50
2	73	2	100
3	72	2	100
4	75	2	100
5	37	1	50
6	02	0	0
7	87	3	150
8	98	4	200
9	10	0	0
10	47	1	50
11	93	4	200
12	21	1	50
13	95	4	200
14	97	4	200
15	69	2	100
Total		31	1550

Estimated average demand $= \frac{31}{15} = 2.07$ PCs

Estimated average revenue $= \frac{1550}{15} = \$103.3$

INVENTORY CONTROL

$$\begin{pmatrix}\text{Total}\\ \text{inventory}\\ \text{cost}\end{pmatrix} = \begin{pmatrix}\text{Purchasing}\\ \text{cost}\end{pmatrix} + \begin{pmatrix}\text{Set up}\\ \text{cost}\end{pmatrix} + \begin{pmatrix}\text{Holding}\\ \text{cost}\end{pmatrix} + \begin{pmatrix}\text{shortage}\\ \text{cost}\end{pmatrix}$$

Demand

- Deterministic: D (fixed no)
  - Stationary
  - Dynamic
- Probabilistic: $X_t \sim f(X_t)$ (RV)
  - Stationary
  - Non-stationary

- **Economic-Order-Quantity (EOQ) Models:-**

Assumption:- The simplest inventory model assumes constant-rate demand with instantaneous order replenishment & no shortage.

Define,

y = Order quantity (number of units)

D = Demand rate (units per unit time)

t_0 = Ordering cycle length (time units)

$$t_0 = \frac{y}{D} \text{ time units}$$

K = Set up cost (ordering cost)

h = Holding cost (dollars per inventory unit per unit time)

Total cost per unit time (TCU) is

$TCU(y)$ = Set up cost per unit time + Holding cost per unit time

$$= \frac{\text{Setup cost} + \text{Holding cost per cycle } t_0}{t_0}$$

$$= \frac{K + h\left(\frac{y}{2}\right) t_0}{t_0}, \quad \text{average inventory level} = \frac{y}{2}$$

$$= \frac{K}{t_0} + \frac{yh}{2}$$

$$= \frac{KD}{y} + \frac{yh}{2}$$

$$\frac{d\,TCU(y)}{dy} = -\frac{KD}{y^2} + \frac{h}{2} = 0$$

$$\Rightarrow y^* = \sqrt{\frac{2KD}{h}}$$

optimum value of $TCU(y) = \frac{KD}{y^*} + \frac{y^* h}{2}$

$$= \sqrt{2hKD}.$$

Ex. 1. $D = 100$ units/day
$K = \$100$ per order
$h = \$0.02$ per unit/day
$L = 0$ days.

$\rightarrow y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\times100\times100}{0.02}} = 1000.$

cycle length, $t_0^* = \frac{y^*}{D} = 10$ days.

$L = 12$
Effective length time
$= L - nt_0$
$= 12 - 1\times10$
$= 2$ day.

Ex. 2. A company orders ground meat at the start of each week to cover the weeks demand of 300lb. The fixed cost per order is \$20. It costs about \$0.03 per lb/day to refrigerate and store the meat (holding cost).

(a) Determine the inventory cost per week of the product ordering policy.

(b). Determine the optimal (inventory policy). The company should use $L = 0$.

$\rightarrow$ $D = 300$ lb/week
$K = \$20$ per order
$h = \$0.03$ per lb/day
$h = \$0.21$ per lb/week
$L = 0$ day

(b) $y = \sqrt{2hKD}$
$= \sqrt{2\times0.21\times20\times300}$
$= 50.199$

(b) Total inventory cost (TIC)
$= K + \frac{D}{2}\times h$
$= 20 + \frac{300}{2}\times0.21$
$= 51.5$

(*) $D \sim N(100, 10^2)$

B = Buffer stock size

$$P\{y^* + B > D\} \leq 0.05$$

$$\Rightarrow P(X_2 \geq B + \mu_2\} \leq 0.05$$

$$\Rightarrow P\left(\frac{X_2 - \mu_L}{\sigma_L} \geq \frac{B}{\sigma_L}\right) \leq 0.05$$

$$\Rightarrow P\left(Z \geq \frac{B}{\sigma_L}\right) \leq 0.05$$

Q. Given the same data for $\alpha = 0.05$ and Daily Demand follows $N(100, 10^2)$; determine the buffer stock size B.

Sol.

μ_L = Average demand during lead time of 2 days

$= 100 \times 2 = 200$

$$\sigma_L = \sqrt{\sigma^2 L} = \sqrt{10^2 \times 2} = 14.14$$

$$P\left(Z \geq \frac{B}{14.14}\right) \leq 0.05$$

$$\rightarrow P\left(Z \leq \frac{B}{14.14}\right) \geq 0.95$$

$$\rightarrow \frac{B}{14.14} = 1.645$$

$$\rightarrow B = 23.2603$$

EOQ with Price Break:-

$$\begin{pmatrix}\text{Total}\\ \text{inventory}\\ \text{Cost}\end{pmatrix} = \begin{pmatrix}\text{cost}\\ \text{per}\\ \text{unit}\end{pmatrix} + \begin{pmatrix}\text{Set}\\ \text{up}\\ \text{cost}\end{pmatrix} + \begin{pmatrix}\text{Holding}\\ \text{cost}\end{pmatrix} + \begin{pmatrix}\text{cost due to not}\\ \text{meeting the}\\ \text{demand}\end{pmatrix}$$

q = price break point

$$TCU(y) = TCU = \frac{KD}{y} + \frac{h}{2}y$$

given cost per unit, $c = \begin{cases} c_1, & y \leq q \\ c_2, & y > q \end{cases}$ $\quad c_1 > c_2$

$$\text{Purchasing cost/unit time} = \frac{c_1 y}{y/D} = Dc_1 \quad \text{when } y \leq q$$

$$= \frac{c_2 y}{y/D} = Dc_2 \quad \text{when } y > q$$

$$TCU_1(y) = Dc_1 + \frac{KD}{y} + \frac{h}{2}y$$

$$TCU_2(y) = Dc_2 + \frac{KD}{y} + \frac{h}{2}y$$

Ex. Lube car specializes in fast automobile oil change. The garage buys car oil in bulk at \$3 per gallon discounted to \$2.50 per gallon if the order quantity is more than 1000 gallons. The garage services approximately 150 cars per day, and each oil change takes 1.25 gallons. Lube car stores buke oil at the cost of \$0.02 per gallon per day. Also cost of placing an order is \$20. There is a 2-day lead time for delivery. Determine the optimal inventory policy.

Sol. The consumption of oil per day is

$D = 150$ cars per day $\times 1.25$ gallons per day $= 187.5$ gallons per day.

$h = \$0.02$ per gallon per day

$K = \$20$ per order

$L = 2$ days

$c_1 = \$3$ per gallon

$c_2 = \$2.50$ per gallon

$q = 1000$ gallons

$$y_m = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \times 20 \times 187.5}{0.02}} = 612.37 \text{ gallons}$$

$$TCU_1(y_m) = C_1 D + \frac{KD}{y_m} + \frac{h y_m}{2}$$

$$= 3 \times 187.5 + \frac{20 \times 187.5}{612.37} + \frac{0.02 \times 612.37}{2}$$

$$= 574.75$$

$$Q^2 + 2\left(\frac{C_2 D - TCU_1(y_m)}{h}\right) Q + \frac{2KD}{h} = 0$$

$$\Rightarrow Q^2 + 2\left(\frac{2.5 \times 187.5 - 574.75}{0.02}\right) Q + \frac{2 \times 20 \times 187.5}{0.02} = 0$$

$$\Rightarrow Q^2 - 10599.74 Q + 375000 = 0$$

$$\Rightarrow Q = 10564.25.$$

Multi-item EOQ with storage Limitation:-

$$\text{Minimize } TCU(y_1, y_2, \dots, y_n) = \sum_{i=1}^{n} \left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2} \right)$$

subject to $\sum_{i=1}^{n} a_i y_i \le A, \quad y_i > 0, \ i=1,2,\dots,n.$

Using Lagrange's optimization technique

$$L(y_1, y_2, \dots, y_n; \lambda) = \sum \left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2} \right) - \lambda \left(\sum a_i y_i - A \right)$$

λ: Lagrange multiplier

$$\frac{\partial L}{\partial y_i} = \left(-\frac{K_i D_i}{y_i^2} + \frac{h_i}{2} \right) - \lambda (a_i) = 0$$

$$\frac{\partial L}{\partial A} = A - \sum a_i y_i = 0$$

Solving the first equation, we get

$$y^* = \sqrt{\frac{2 K_i D_i}{h_i - 2\lambda a_i}}.$$

Q.

Item i	K_i (\$)	D_i (units per day)	h_i (\$)	a_i (ft^2)
1	10	2	0.30	1
2	5	4	0.10	1
3	15	4	0.20	1

The unconstrained optimum values $y_i^* = \sqrt{\frac{2K_iD_i}{h_i}}$, $i=1,2,3$ are $y_1^* = 11.55$, $y_2^* = 20$, $y_3^* = 24.49$ units which violate $y_1 + y_2 + y_3 \le 25$.

Construct the following table:-

λ	y_1	y_2	y_3	$\sum a_i y_i - A$
−0.1	9	11.5	14.9	10.4
−0.25	7.1	8.2	11.3	1.6
−0.3	6.7	7.6	10.6	−0.1

Probabilistic EOQ Model:-

$f(x)$ = pdf of demand, x, during lead time
D = Expected demand per unit time
h = Holding cost per inventory unit per unit time
p = Shortage cost per inventory unit
K = Set up cost per order
R = Reorder point

1. Set up: The approximate number of orders per unit time is $\frac{D}{y}$,
   The set up cost per unit time $= \frac{KD}{y}$.

2. Expected holding cost:- I = average inventory level
   hI = expected holding cost per unit time

Inventory today $= y + E(R-x)$
Inventory at the end of the cycle $= E(R-x)$.

$$\text{Average inventory level} = I = \frac{(y + E(R-x)) + E(R-x)}{2}$$

$$= \frac{y}{2} + R - E(x)$$

3. Expected shortage cost:- $x > R$

Expected value per cycle, $S = \int_R^\infty (x-R) f(x)\,dx$

p : proportional to the shortage quantity
pS : expected shortage cost per cycle

$$\text{shortage cost per unit time} = \frac{pS}{y/D} = \frac{pDS}{y}.$$

Total cost function per unit time,

$$TCU(y,R) = \frac{DK}{y} + h\left(\frac{y}{2} + R - E(x)\right) + \frac{pD}{y}\int_R^\infty (x-R) f(x)\,dx$$

The optimal values y^*, R^* are determined from

$$\frac{\partial TCU}{\partial y} = -\left(\frac{DK}{y^2}\right) + \frac{h}{2} - \frac{pDS}{y^2} = 0$$

$$\frac{\partial TCU}{\partial R} = h - \left(\frac{pD}{y}\right)\int_R^\infty f(x)\,dx = 0$$

$$\Rightarrow y^* = \sqrt{\frac{2D(K+pS)}{h}}$$

$$\int_R^\infty f(x)\,dx = \frac{hy^*}{pD}$$

for $R = 0$,

$$\hat{y} = \sqrt{\frac{2D(K+pE(x))}{h}}$$

$$\tilde{y} = \frac{pD}{h}$$

Unique optimal values of y and R exist when $\tilde{y} \geq \hat{y}$.
The smallest value of $y^* = \sqrt{\frac{2KD}{h}}$ occurs when $S=0$.

Ex. Electro uses resin in its manufacturing process at the rate of 1000 gallons per month. It cost Electro \$100 to place an order. The holding cost per gallon per month is \$2, and the shortage cost per gallon is \$10. Historical data shows that the demand during lead time is uniform in the range (0, 100) gallons. Determine the optimal ordering policy for Electro.

Solution:- $D = 1000$ gallons per month
$K = \$100$ per order
$h = \$2$ per gallon per month
$p = \$10$ gallon
$f(x) = \frac{1}{100}$, $0 \leq x \leq 100$
$E[x] = 50$ gallons

$$\hat{y} = \sqrt{\frac{2 \times 1000 (100 + 10 \times 50)}{2}} = 774.6 \text{ gallons}$$

$$\tilde{y} = \frac{10 \times 1000}{2} = 5000 \text{ gallons}$$

$\tilde{y} \geq \hat{y}$, a unique solution exists for y^* and R^*.

$$S = \int_R^{100} (x - R) \frac{1}{100} dx = \frac{R^2}{200} - R + 50$$

$$y_i = \sqrt{\frac{2 \times 1000 (100 + 10S)}{2}} = \sqrt{100000 + 10000S} \text{ gallons}$$

$$\int_R^{100} \frac{1}{100} dx = \frac{2y_i}{10 \times 1000}$$

$$R_i = 100 - \frac{y_i}{50}.$$

Iteration 1:-

$$y_1 = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 \times 1000 \times 100}{2}} = 316.23 \text{ gallons}$$

$$R_1 = 100 - \frac{316.23}{50} = 93.68 \text{ gallons}$$

Iteration 2:-

$$S = \frac{R_1^2}{100} - R_1 + 50 = .19971 \text{ gallons}$$

$$y_2 = \sqrt{100,000 + 10000 \times .19971} = 319.37 \text{ gallons}$$

Hence, $R_2 = 100 - \frac{319.39}{50}$

$= 93.612$

Iteration 3:-

$$S = \frac{R_2^2}{200} - R_2 + 50 = .20399 \text{ gallons}$$

$$y_3 = \sqrt{100000 + 10000 \times .20399}$$

$= 319.44$ gallons

thus, $R_3 = 100 - \frac{319.44}{50} = 93.611$ gallons

$R^* = 93.611$ gallons, $y^* = 319.44$ gallons.

- Single Period Model:-

☒ News-Boy Problem:- $D \to$ Demand
$y \to$ stock

$K =$ Set up cost
$h =$ Holding cost per held unit during the period
$p =$ Penalty cost per shortage unit during the period
$y =$ order quantity
$x =$ Inventory on hand before an order is placed.

Assumptions:- 1. Demand occurs instantaneously at the start of the period and immediately after order is received.
2. No set up cost.

First assume that demand D is continuous.
Expected cost for the period.

$$E\{c(y)\} = h\int_0^y (y-D)f(D)\,dD + p\int_y^\infty (D-y)f(D)\,dD$$

$$\frac{\partial E\{c(y)\}}{dy} = h\int_0^y f(D)\,dD - p\left[1-\int_0^y f(D)\,dD\right]$$

$$\therefore P(D \le y^*) = \frac{p}{p+h}.$$

Ex. The buying price of a newspaper is 30; selling price is 75. Newspaper unsold has a resale value (Scrap value) = 5. How many copies of newspaper should be purchased to maximize profit, when
(a) Demand is $D \sim N(300, 20^2)$
(b) Demand is discrete pmf.

D	200	220	300	320	340
f(D)	·1	·2	·4	·2	·1

Solution:- (a) penalty, $p = 75 - 30 = 45$
$h = 30 - 5 = 25$

$$P(D \le y^*) = \frac{45}{45+25} = \frac{45}{70}$$

$\Rightarrow P\left(z \le \frac{y^* - 300}{20} \right) = 0.6428$

$\Rightarrow \frac{y^* - 300}{20} = 0.37$

$\Rightarrow y^* = 307.8$

The optimal order is approximately 308 copies.

(b)

y	200	220	300	320	340
$P(D \le y)$	.1	.3	.7	.9	1.0

$P(D \le 220) \le .643 \le P(D \le 300)$

$\therefore y^* = 300$ copies.

Ex. The daily demand for an item during a single period occurs instantaneously at the start of the period. The pdf of demand is $U(0, 10)$. The unit holding cost of the item during the period is \$.50, and the unit penalty cost running out of stock is \$4.50. A fixed cost of \$25 is incurred each time an order is placed. Determine the optimal inventory policy for the item?

Sol. $\frac{p}{p+h} = \frac{4.5}{4.5 + 0.5} = .9$

$$P(D \le y^*) = \int_0^{y^*} \frac{1}{10}\, dD = \frac{y^*}{10}$$

$S = y^* = 9.$

The expected cost function is

$$E(C(y)) = 0.5 \int_0^{y} \frac{1}{10}(y - D)\, dD + 4.5 \int_y^{10} \frac{1}{10}(D - y)\, dD$$

$$= 0.25y^2 - 4.5y + 22.5$$

$E[C(s)] = K + E\{C(S)\}$

$\Rightarrow 0.25s^2 - 4.5s + 22.5 = 25 + 0.25S^2 - 4.5S + 22.5$

$\Rightarrow s^2 - 18s - 19 = 0$ (Put $S = 9$)

$\Rightarrow s = -1$ or 19

Since $s < S$ but $s = -1$, so s has no feasible value.

GOAL PROGRAMMING

Let $(x_1, x_2, \ldots, x_n)$ be the vector of decision variable.
Let $Z_1(\underset{\sim}{x}), Z_2(\underset{\sim}{x}), \ldots, Z_K(\underset{\sim}{x})$ denote 'K' objective functions to be optimized simultaneously.

$$\text{Min } Z_1(x), Z_2(x), \ldots, Z_K(x)$$

simultaneously

$$\text{s.t.} \quad Ax = b$$
$$Dx \geq d$$
$$x \geq 0$$

A feasible solution $\bar{x}$ is said to be Pareto optimal solution (or vector minimum, non-dominated solution, equilibrium solution, etc) to the Multi[redacted]-Objective Linear Programming, if there exists no other feasible solution x that is better than $\bar{x}$ for every objective function & strictly better for at least one objective function.

i.e., there exists no feasible solution 'x'

$$Z_r(x) \leq Z_r(\bar{x}) \quad , r = 1, 2, \ldots, K$$
$$Z_r(x) < Z_r(\bar{x}) \text{ for at least one } r.$$

Dominated solution is never a desirable solution to implement. Because there are other solutions better than this dominated solution for every objective function.

So for a feasible solution to be a candidate to be considered for a MOLP is must be undominated.

Very efficient algorithm has been developed for enumerating the set of all non-dominated solution to MOLP. this set is commonly known as the efficient frontier.

Practical Approaches for handling MOLP in current use :—

Suppose the decision makers determined that c_1 units of Z_1 (in whatever units this objective function is measured) is equivalent to or has the same merit or value to c_2 units of Z_2 of its own units that is equivalent to c_3 units of Z_3 of its own units & so on.

This vector $(c_1, c_2, \ldots, c_K)$ gives a complete compromise or exchange information between various objective functions & so can be called as exchange vector.

So, MOLP can be written as

$$\text{Min} \quad \frac{Z_1(x)}{C_1} + \frac{Z_2(x)}{C_2} + \cdots + \frac{Z_K(x)}{C_K}$$

s.t. $Ax = b$
$Dx \geq d$
$x \geq 0$

Let $W_1, W_2, \ldots, W_K$ be the set of weight vectors

$$\text{Min } Z(x) = \sum_{r=1}^{K} W_r Z_r(x)$$

Such that $Ax = b$
$Dx \geq d$
$x \geq 0$

If each $w_r \geq 0$, then every optimum solution is a non dominated solution.

Ques:- Consider the fertilizer makers product mixed who used High pH and Low pH fertilizer. They require Raw material RM_1, RM_2, RM_3 with the following data.

Item	Ton. reqd. to make one ton of		
	Hi-pH	Lo-pH	Available
RM1	2	1	1500
RM2	1	1	1200
RM3	1	0	500
Net profit	15	10	

→ Obj. function is $Z_1(x) = 15x_1 + 10x_2$
$Z_2(x) = 222x_1 + 107x_2$
$Z_3(x) = 222x_1$

$$\text{Max } \{Z_1(x), Z_2(x), Z_3(x)\}$$

s.t. $2x_1 + x_2 \leq 1500$
$x_1 + x_2 \leq 1200$
$x_1 \leq 500$
$x_1, x_2 \geq 0$

Suppose the decision makers are decided weight for the objective function are 0.5, 0.25, 0.25, then we take a compromise solution to this MOP to be an optimal solution.

$$\text{Max} \quad 0.5(15x_1+10x_2) + 0.25(222x_1+107x_2) + 0.25(222x_1)$$

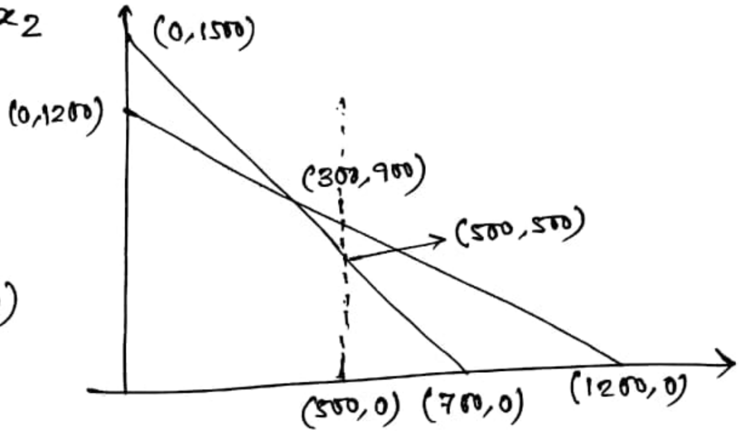
i.e. Max $118.5x_1 + 31.75x_2$

s.t. $2x_1 + x_2 \leq 1500$

$x_1 + x_2 \leq 1200$

$x_1 \leq 500$

$x_1, x_2 \geq 0$



<u>Sol.</u> $(x_1, x_2) \equiv (500, 500)$

<u>Goal Programming</u> :- Goal programming may be used to solve linear programs with multiple objectives, with each objective viewed as a 'goal'.

- d_i^+ and d_i^-, deviation variables, are the amounts a targeted goal i is overachieved or underachieved, respectively.
- The goals themselves are added to the constraint set with d_i^+ and d_i^- acting as the surplus and slack variables.
- One approach to goal programming is to satisfy goals in a priority sequence. Second-priority goals are pursued without reducing first priority goals, etc.
- For each priority level, the objective function is to minimize the (weighted) sum of goal deviations.
- Previous "optimal" achievements of goals are added to the constraint set so that they are not degraded while trying to achieve lesser priority goals.

Goal Programming (Multiple Objectives/Goals)

Q.1.

	A	B	Availability
I	3	4	50
II	5	6	60
III	1		
Profit	70	80	

Profit to be kept above 800

Quantity of item III to be ordered is 8 or less.

$$\text{Max } 70x_1 + 80x_2$$

s.t. $3x_1 + 4x_2 \le 50$

$5x_1 + 6x_2 \le 60$

$x_1 \le 8$

→ Minimize total production so that transportation is easy.

1. Rigid constraints: $3x_1 + 4x_2 \le 50$ ⟶ $\le$: min ρ_1
   $5x_1 + 6x_2 \le 60$

2. $70x_1 + 80x_2 \geq 800$ ⟶ $\geq$: max η.

3. $x_1 \le 8$

4. $x_1 + x_2 \le 8$

Minimize $[(\rho_1 + \rho_2), \eta_3, \rho_4, \rho_5]$

s.t. $\left.\begin{array}{l} 3x_1 + 4x_2 - \eta_1 - \rho_1 = 50 \\ 5x_1 + 6x_2 + \eta_2 - \rho_2 = 60 \end{array}\right\}$ Min $\rho_1 + \rho_2$.

$70x_1 + 80x_2 + \eta_3 - \rho_3 = 800$ min η_3

$x_1 + \cancel{x_2} + \eta_4 - \rho_4 = 8$ min ρ_4

$x_1 + x_2 + \eta_5 - \rho_5 = 8$ Min ρ_5

$x, \eta, \rho \geq 0$

Integer Programming (Formulation)

Q. I have been approached by 3 telephone companies to subscribe to their long-distance service in the US. MaBell will charge a flat \$16 per month plus \$.25 a minute. PaBell will charge \$25 a month but will reduce the per-minute cost to \$21. As for BabyBell, the flat monthly charge is \$18, and cost perminute is \$.22. I usually make an average of 200 minutes of long distance calls a month. Assuming that I don't pay the flat monthly fee unless I make calls and that I can apportion my calls among all 3 companies as I please, how should I use the 3 companies to ~~optimi~~ minimize my monthly telephone bill?

Sol.

x_1 = MaBell long-distance minutes per month

x_2 = PaBell long-distance minutes per month

x_3 = BabyBell long-distance minutes per month

$y_1 = 1$ if $x_1 > 0$ and 0 if $x_1 = 0$

$y_2 = 1$ if $x_2 > 0$ and 0 if $x_2 = 0$

$y_3 = 1$ if $x_3 > 0$ and 0 if $x_3 = 0$

$$x_j \le M y_j, \quad j = 1,2,3$$

$x_j \le 200$ for all j, since I make about 200 min. of calls a month, $M = 200$.

The complete model is

$$\text{Minimize } Z = 0.25x_1 + 0.21x_2 + 0.22x_3 + 16y_1 + 25y_2 + 18y_3$$

s.t.

$$x_1 + x_2 + x_3 = 200$$

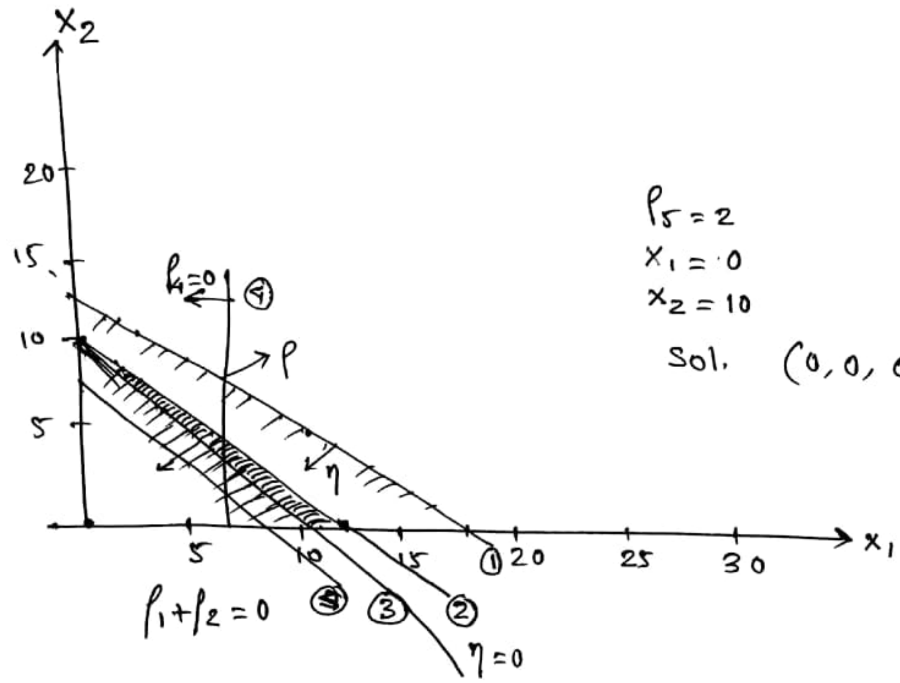
$$x_1 \le 200y_1$$

$$x_2 \le 200y_2$$

$$x_3 \le 200y_3$$

$$x_1, x_2, x_3 \ge 0$$

$$y_1, y_2, y_3 = (0, 1).$$



$\rho_5 = 2$
$x_1 = 0$
$x_2 = 10$
Sol. $(0, 0, 0, 2)$

Q.2. A and B assembled

A $\downarrow$ 4hrs, B $\downarrow$ 3hrs

RT = 90hrs

Profit 7 & 8

OT is available but reduces the profit by 1.

Demand 30 & 40

Goals are:-
1. Targets have to be made & RT = 90
2. Minimize OT
3. Maximize profit

$x, \eta, \rho \geq 0$

x_1, x_2 of A in RT & OT
y_1, y_2 of B in RT & OT

$$\text{Min}\left[\left(\eta_1 + \eta_2 + \rho_3\right), \rho_4, \eta_5\right]$$

$x_1 + x_2 + \eta_1 - \rho_1 = 30$ $\quad$ Min $\eta_1 + \eta_2 + \rho_3$

$y_1 + y_2 + \eta_2 - \rho_2 = 40$

$4x_1 + 3y_1 + \eta_3 - \rho_3 = 90$

$4x_2 + 3y_2 + \eta_4 - \rho_4 = 20$ $\quad$ Min ρ_4

$7x_1 + 6x_2 + 8y_1 + 7y_2 + \eta_5 - \rho_5 = 200$ $\quad$ Min η_5

Nonlinear Programming

Convex Sets

The concept of convexity is of great importance in the study of optimization problems. Convex sets, polyhedral sets, and separation of disjoint convex sets are used frequently in the analysis of mathematical programming problems, the characterization of their optimal solutions, and the development of computational procedures.

Under convex sets, we shall study

- Convex Hulls
- Closure and interior of a convex set
- Separation and support of convex sets
- Convex cones and polarity
- Polyhedral sets, extreme points,
  and extreme directions
- Linear programming and the simplex method

Definition 1.1 (Convex Set). A nonempty set S in R^n is said to be convex if the line segment joining any two points in the set is contained in S. Equivalently, if $x, y \in S$ and $\lambda \in [0, 1]$, then $\lambda x + (1 - \lambda)y \in S$.

Examples of Convex Sets:

1. $S = \{(x, y, z) : x + 2y - z = 4\} \subset R^3$
2. $S = \{(x, y, z) : x + 2y - z \leq 4\} \subset R^3$
3. $S = \{(x, y, z) : x + 2y - z \leq 4, 2x - y + z \leq 6\}$
4. $S = \{(x, y) : y \geq |x|\} \subset R^2$
5. $S = \{(x, y) : x^2 + y^2 \leq 4\} \subset R^2$

6. S is the set of solutions to the linear programming problem: Minimize $c^t x$ subject to $Ax = b,\ x \geq 0$.

Definition 1.2. Let $x_1, x_2, \ldots, x_k$ be in R^n. Then $x = \sum_{i=1}^{k} \lambda_i x_i$ is called a **convex combination** of the k points provided $\lambda_i \geq 0$ for each i and $\sum_{i=1}^{k} \lambda_i = 1$. If the nonnegativity condition on λ_is is dropped, then the combination is called an **affine combination**.

Exercise 1.1. Shaw that a set S is convex if, and only if, for any positive integer k, any convex combination any k points in S is in S.

Lemma 1.1. If S and T are convex sets, then

1. $S \cap T$ is convex,
2. $S + T = \{x + y : x \in S,\ y \in T\}$ is convex,
3. $S - T = \{x - y : x \in S,\ y \in T\}$ is convex.

Definition 1.3(Convex Hull). Let S be a nonempty set in R^n. The convex hull of S is defined as the set of all convex combinations of S. Convex hull of S is denoted by $H(S)$. Note that $x \in H(S)$ if, and only if, there exist $x_1, x_2, \ldots, x_k$ (k is a positive integer) in S such that x is convex combination of $x_1, x_2, \ldots, x_k$.

Exercise 1.2. What is the convex hull of three noncollinear points in R^2?

Definition 1.4(Polytope and Simplex). The convex hull of a finite number of points $x_1, x_2, \ldots, x_{k+1}$ in R^n is called a **polytope**. The polytope is called a simplex with vertices $x_1, x_2, \ldots, x_{k+1}$ provided $x_{k+1} - x_1, x_{k+1} - x_2, \ldots, x_{k+1} - x_k$ are linearly independent.

Exercise 1.3. Is the polytope formed by $(1,1), (2,2), (0,0)$ a simplex? Draw a simplex in R^2. Can you draw a simplex in R^2 with four vertices in R^2?

Theorem 1.1 (Caratheodory). Let $S \subseteq R^n$. If $x \in H(S)$, then x can be written as a convex combination of $n + 1$ points from S.

Proof. Let $x \in H(S)$. By definition, there exist $k+1$ points $x_1, x_2, \ldots, x_{k+1}$ in S (k is a nonnegative integer) such that $x = \sum_{i=1}^{k+1} \lambda_i x_i$ with $\lambda_i > 0$ for all i. If $k \leq n$, there is nothing to prove. Suppose $k \geq n+1$. It follows $x_{k+1} - x_1, x_{k+1} - x_2, \ldots, x_{k+1} - x_k$ are linearly dependent. This implies $x = \sum_{i=1}^{k} \mu_i (x_{k+1} - x_i) = 0$ for some real numbers $\mu_1, \mu_2, \ldots, \mu_k$, at least one of them different from zero. Let $\mu_{k+1} = -\sum_{i=1}^{k} \mu_i$. Then $\sum_{i=1}^{k+1} \mu_i = 0$ and $\sum_{i=1}^{k+1} \mu_i x_i = 0$. Since λ_is are positive, we can choose a constant c such that $\beta_i = \lambda_i - c\mu_i \geq 0$ for all i and equal to zero for at least one i. Note that

$$x = \sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^{k+1} \lambda_i x_i - c \sum_{i=1}^{k+1} \mu_i x_i = \sum_{i=1}^{k+1} \beta_i x_i.$$

From the above, it follows that x can be written as a convex combination of k points from S. If $k \leq n$, the proof is complete. Otherwise we repeat the argument to show that x can be written as a convex combination of fewer than k points from S. This argument can be continued until x can be written as a convex combination of $n+1$ points from S. $\square$

Definition 1.5. Let $S \subseteq R^n$. A point $x \in R^n$ is said to be an interior point of S if $N_\epsilon(x) \subseteq S$ for some $\epsilon > 0$, where $N_\epsilon(x) = \{u \in R^n : ||u - x|| < \epsilon\}$. A point x is said to be in the closure of S if there exists a sequence $x_1, x_2, \ldots$ of points from S such that the sequence converges to x. The set of all interior points of S is denoted by $int(S)$ and the set of all points in the closure of S is denoted by $Cl(S)$. The set of all points in the intersection of $Cl(S)$ and $Cl(S^c)$ is called the boundary of S and is denoted by ∂S (S^c stands for complement of S in R^n). A set S is said to be open if $int(S) = S$; and S is said to be closed if $Cl(S) = S$. A set S is said to be compact if it is closed and bounded.

Exercise 1.4. Show that every polytope in R^n is a compact convex set.

Theorem 1.2. Let S be a nonempty convex set in R^n with nonempty interior. Let $x_1 \in Cl(S)$ and let $x_2 \in int(S)$. Then $\lambda x_1 + (1-\lambda)x_2 \in int(S)$ for all $\lambda \in (0,1)$.

Proof. Since $x_2 \in int(S)$, there exists an $\epsilon > 0$ such that $N_\epsilon(x_2) \subset S$. Fix any $\lambda \in (0,1)$ and let $y = \lambda x_1 + (1-\lambda)x_2$. Will show that $y \in int(S)$ by showing that $N_\delta(y) \subset S$ where $\delta = (1-\lambda)\epsilon$.

Let $z \in N_\delta(y)$. Since $x_1 \in Cl(S)$, there exists a $z_1 \in S$ such that $||z_1 - x_1|| < \frac{\delta - ||z-y||}{\lambda}$.

Let $z_2 = \frac{z-\lambda z_1}{1-\lambda}$. Then

$$\begin{aligned} ||z_2 - x_2|| &= ||\frac{z - \lambda z_1}{1-\lambda} - x_2|| \\ &= ||\frac{(z-\lambda z_1) - (y - \lambda x_1)}{1-\lambda}|| \\ &= \frac{1}{1-\lambda}||(z-y) + \lambda(x_1 - z_1)|| \\ &\leq \frac{1}{1-\lambda}[||(z-y)|| + \lambda||(x_1 - z_1)||] \\ &< \epsilon \end{aligned}$$

Corollary 1.1. If S is a convex set, then so is $int(S)$.

Corollary 1.2. If S is a convex set with nonempty interior, then $Cl(S)$ is a convex set.

Corollary 1.3. If S is a convex set with nonempty interior, then $Cl(int(S)) = Cl(S)$.

Corollary 1.4. If S is a convex set with nonempty interior, then $int(Cl(S)) = int(S)$.

Separation and Support of Convex Sets

The notion of separation and support of disjoint convex sets plays an important role in deriving results regarding optimality conditions and also in computing optimal solutions.

Theorem 1.3. Let S be a nonempty closed convex set in R^n and let $y \in R^n \setminus S$. Then, there exists a unique point $\bar{x} \in S$ with minimum distance from S. Furthermore, $\bar{x}$ is the minimizing point if, and only if, $(x - \bar{x})^t(\bar{x} - y) \geq 0$ for all $x \in S$.

Proof. Let $\gamma = \inf_{x \in S} ||x - y||$. Since S is closed and $y \notin S$, $\gamma > 0$. It follows that there exists a sequence $x_k \in S$ such that $||x_k - y|| \to \gamma$. Clearly x_k is a bounded sequence and hence must have a convergent subsequence. Without loss of generality, we may assume that x_k itself converges to a point $\bar{x}$. Since S is closed, $\bar{x} \in S$ and $||\bar{x} - y|| = \gamma$.

Suppose there is another point $Z \in S$ such that $||z - y|| = \gamma$. Since S is convex, $\frac{1}{2}(z + \bar{x}) \in S$ and

$$\gamma \leq ||(y - \frac{z + \bar{x}}{2})|| \leq \frac{1}{2}(||y - z|| + ||y - \bar{x}||) = \gamma.$$

It follows that $y - z = p(y - \bar{x})$. Since $||y - z|| = ||y - \bar{x}|| = \gamma$, it follows that either $p = 1$ or $p = -1$. If $p = -1$, then we must have $y = \frac{z+\bar{x}}{2}$ which in turn implies that

$y \in S$ which is a contradiction. It follows that $p = 1$ and hence $z = \bar{x}$.

Next, suppose $\bar{x}$ is such that $x - \bar{x})^t(\bar{x} - y) \geq 0$ for all $x \in S$. Let $x \in S$. Then,

$$\begin{aligned} ||x-y||^2 &= ||x-\bar{x}+\bar{x}-y)||^2 \\ &= ||x-\bar{x}||^2 + ||\bar{x}-y||^2 + 2(x-\bar{x})^t(\bar{x}-y) \end{aligned}$$

Since $||x - \bar{x}||^2 \geq 0$ and $x - \bar{x})^t(\bar{x} - y) \geq 0$, it follows from the above that $||x - y||^2 \geq ||\bar{x} - y||^2$ and hence the $\bar{x}$ is the distance minimizing point.

Next, assume that $\bar{x}$ is the distance minimizing point. Fix $x \in S$ and $\lambda \in (0, 1)$. Since S is convex, $\bar{x} + \lambda(x - \bar{x}) \in S$ and since $\bar{x}$ is the minimizing point

$$||y - \bar{x} - \lambda(x - \bar{x})||^2 \geq ||y - \bar{x}||^2$$

Also

$$||y - \bar{x} - \lambda(x - \bar{x})||^2 = ||y - \bar{x}||^2 + \lambda^2||x - \bar{x}||^2 + 2\lambda(\bar{x} - y)^t(x - \bar{x})$$

Rearranging and noting that $\bar{x}$ is the minimizing point, we get

$$\lambda^2||x - \bar{x}||^2 + 2\lambda(\bar{x} - y)^t(x - \bar{x}) = ||y - \bar{x} - \lambda(x - \bar{x})||^2 - ||y - \bar{x}||^2 \geq 0$$

Dividing both sides by λ and taking limit as $\lambda \to 0$, we get $(x - \bar{x})^t(\bar{x} - y) \geq 0$ and this completes the proof. $\square$

Definition 1.6 (Hyperplane). A *hyperplane* in R^n is a set of the form $H = \{x : p^t x = \alpha\}$, where p is a nonzero vector in R^n and α is a real number. The sets

$$H^+ = \{x : p^t x \geq \alpha\} \text{ and } H^- = \{x : p^t x \leq \alpha\}$$

are called closed half spaces, and the sets

$$H^+ = \{x : p^t x > \alpha\} \text{ and } H^- = \{x : p^t x < \alpha\}$$

are called open half spaces.

The vector p is called the **normal** to the hyperplane H.

A hyperplane passing through a point $\bar{x}$ can be written as $H = \{x : p^t(x - \bar{x}) = 0\}$.

Definition 1.7 (Separation). Let S and T be two nonempty subsets of R^n. A hyperplane $H = \{x : p^t x = \alpha\}$ is said to **separate** S and T if $p^t x \geq \alpha$ for all $x \in S$, and $p^t x \leq \alpha$ for all $x \in T$. The separation is said to be **proper** if $SUT \not\subseteq H$. The separation is said to be **strict** if $p^t x > \alpha$ for all $x \in S$, and $p^t x < \alpha$ for all $x \in T$. The separation is said to be strong if there exists an $\epsilon > 0$ such that $p^t x > \alpha + \epsilon$ for all $x \in S$, and $p^t x \leq \alpha$ for all $x \in T$.

Theorem 1.4. Let S be a nonempty closed convex set in R^n and let $y \in R^n \setminus S$. Then there exist a nonzero vector p and a real number of α such that for all $x \in S$, $p^t y > \alpha \geq p^t x$.

Proof. From Theorem 1.3, there exists a unique $\bar{x} \in S$ such that

$$(x - \bar{x})^t (\bar{x} - y) \geq 0 \text{ for all } x \in S.$$

Taking $p = y - \bar{x}$ and $\alpha = p^t \bar{x}$, we have $p \neq 0$ and $p^t x \leq \alpha$ for all $x \in S$. Also, $0 < p^t p = p^t (y - \bar{x}) = p^t y - p^t \bar{x} = p^t y - \alpha$ which implies $p^t y > \alpha$. $\square$

Corollary 1.5. Every closed convex set in R^n is the intersection of all halfspaces containing it.

Proof. Let S be a closed convex set (nonempty) in R^n. Suffices to show that if y is in the intersection of half spaces containing S, then $y \in S$. Suppose, to the contrary, $y \notin S$. From the theorem, there exist a nonzero vector p and a real number of α such that for all $x \in S$, $p^t y > \alpha \geq p^t x$. Note that the halfspace $H = \{x : p^t x \leq \alpha\}$ contains S. Since y is in the intersection of halfspaces containing S, y must belong to this halfspace. But this implies that $p^t y \leq \alpha$ which is a contradiction. $\square$

Corollary 1.6. Let S be a nonempty set in R^n and let $y \notin Cl(H(S))$. Then, there exists a hyperplane that strongly separates S and y.

Proof. Exercise.

Theorem 1.5 (Farka's Lemma). Let A be an $m \times n$ real matrix and let $c \in R^n$. Then, exactly one of the following systems has a solution:

System 1. $Ax \leq 0$ and $c^t x > 0$ for some $x \in R^n$

System 2. $A^t y = c$ and $y \geq 0$ for some $y \in R^m$

Proof. If both systems have solutions, say, x and y, then $0 < c^t x = y^t Ax \leq 0$. The last inequality follows from $Ax \leq 0$ and $y \geq 0$. From this contradiction, it follows that both systems cannot have solutions simultaneously.

To complete the proof, assume that System 2 has no solution. Let $S = \{u : u = A^t y,$ for some $y \geq 0\}$. Since System 2 has no solution, $c \notin S$. From Theorem 4, there exists a nonzero x and a number α such that $x^t u \leq \alpha$ for all $u \in S$ and $x^t c > \alpha$. Since $0 \in S$, $\alpha \geq 0$. It follows that $x^t A^t y \leq 0$ for all $y \geq 0$, and hence $x^t A^t \leq 0$. Thus, we have x satisfying $Ax \leq 0$ and $c^t x > 0$. □

Corollary 1.7. (Gordan's Theorem). Let A be an $m \times n$ real matrix. Then, exactly one of the following systems has a solution:

System 1. $Ax < 0$ for some $x \in R^n$

System 2. $A^t y = 0$ and $y \geq 0$ for some nonzero $y \in R^m$

Proof. Exercise. Deduce the proof from Farka's lemma.

Ex. Prove Farka's Lemma from Gordan's theorem.

Proof. Apply Gordan's theorem to the matrix $M = \begin{bmatrix} A & -e \\ -c^t & 1 \end{bmatrix}$. Note that $M(x, s)^t < 0$ is equivalent to system I of Farka's Lemma. Complete the proof.

Corollary 1.8. Let A be an $m \times n$ real matrix and let $c \in R^n$. Then, exactly one of the following systems has a solution:

System 1. $Ax \leq 0,\ x \geq 0,\ c^t x > 0$ for some $x \in R^n$

System 2. $A^t y \geq c$ and $y \geq 0$ for some $y \in R^m$

Proof. Exercise (Use $[A^t - I]$ in System 2).

Corollary 1.9. Let A be an $m \times n$ matrix, let B be an $l \times n$ matrix and let $c \in R^n$. Then, exactly one of the following systems has a solution:

System 1. $Ax \leq 0,\ Bx = 0,\ c^t x > 0$ for some $x \in R^n$

System 2. $A^t y + B^t z = c$ and $y \geq 0$ for some $y \in R^m$ and $z \in R^l$.

Proof. Exercise (Use $[A^t B^t \; - \; B^t]$ for A and $z = z^+ - z^-$ in System 2).

Support of Sets at Boundary Points

Definition 1.8. Let S be a nonempty set in R^n, and let $\bar{x} \in \partial S$. A hyperplane $H = \{x : p^t(x - \bar{x}) = 0\}$ is called a supporting hyperplane of S at $\bar{x}$ if either $S \subseteq H^+$ or $S \subseteq H^-$. If, in addition, $S \not\subseteq H$, then H is called a proper supporting hyperplane of S at $\bar{x}$.
See Figure-6.

Exercise 5. Show that the hyperplane $H = \{x : p^t(x - \bar{x}) = 0\}$ is a supporting hyperplane of S at $\bar{x} \in \partial S$ if, and only if, either $p^t\bar{x} = \inf\{p^t x : x \in S\}$ or $p^t\bar{x} = \sup\{p^t x : x \in S\}$.

It will be shown that convex sets have supporting hyperplanes at each of the boundary points.

Theorem 1.6. Let S be a nonempty convex set in R^n, and let $\bar{x} \in \partial S$. Then, there exists a supporting hyperplane $H = \{x : p^t(x - \bar{x}) = 0\}$ that supports S at $\bar{x}$.

Proof. Since $\bar{x} \in \partial S$, there is a sequence $y_k \notin Cl(S)$ such that $y_k \to \bar{x}$. Since, $Cl(S)$ is a closed convex set, for each k, there exists a nonzero p_k such that $p_k^t x \leq y_k^t \bar{x}$ for all $x \in S$. We may assume, without loss of generality, that $||p_k|| = 1$ for each k and that p_k is a convergent sequence. Let p be the limit of p_k. Note that $p \neq 0$ as $||p||$ will be 1. Taking limits in the above inequality, we get $p^t x \leq p^t \bar{x}$. Therefore, $H = \{x : p^t(x - \bar{x}) = 0\}$ supports S at $\bar{x}$. $\square$

Exercises

E1.6. Let S be a nonempty convex set in R^n. If $\bar{x} \notin int(S)$, then show that there exists a nonzero p such that $p^t(x - \bar{x}) \leq 0$ for all $x \in S$.

E1.7. Let S be a nonempty set in R^n. If $y \notin Cl(H(S))$, then show that there exists a hyperplane that separates y and S.

E1.8. Let S be a nonempty set in R^n, and let $\bar{x} \in \partial S \cap \partial H(S)$. Show that there exists a hyperplane that supports S at $\bar{x}$.

Separation of Two Convex Sets

Theorem 1.7. Let S and T be two nonempty disjoint convex sets in R^n. Then there exists a hyperplane that separates S and T, that is, there exists a nonzero p such that

$$p^t x \leq p^t y \qquad \text{for all } x \in S \text{ and for all } y \in T.$$

Proof. Let $U = S - T$. Then U is a convex set and $0 \notin U$ as S and T are disjoint. In particular, $0 \notin int(U)$. Therefore, there exists a nonzero p such that $p^t u \leq 0$ for all $u \in U$. That is, $p^t(x - y) \leq 0$ for all $x \in S$ and for all $y \in T$. Hence the theorem. $\Box$

Exercises

E1.9. How will you define proper, strict and strong separation of sets?

E1.10. Let S and T be two nonempty convex sets in R^n with $int(T) \neq \emptyset$ and $S \cap int(T) = \emptyset$. Show that there exists a hyperplane that separates S and T.

E1.11. Deduce Gordan's theorem using results on separation of convex sets.

E1.12. Let S and T be two nonempty sets in R^n with their convex hulls having nonempty interiors. Assume that $H(S) \cap int(H(T)) = \emptyset$. Show that there exists a hyperplane that separates S and T.

Strong Separation of Convex Sets

Theorem 1.8. Let S and T be two nonempty disjoint closed convex sets in R^n. If S is bounded, then there exists a hyperplane that strongly separates S and T, that is, there exists a nonzero p and an $\epsilon > 0$ such that

$$p^t x \geq \epsilon + p^t y \qquad \text{for all } x \in S \text{ and for all } y \in T.$$

Proof. Let $U = S - T$. Then U is a convex set and $0 \notin U$ as S and T are disjoint. Note that U is a closed set. To see this, let $u^k \in U$ be a sequence converging to u. Then, for each k, there exist $x^k \in S$ and $y^k \in T$ such that $u^k = x^k - y^k$. Since S is compact, we may assume, without loss of generality, that x^k converges to some $x \in S$ (as S is closed). This implies that $y^k \to x - u$. Since T is closed $y \in T$. Hence $u = x - y$ and

$u \in U$. From Theorem 1.4, exists a nonzero p and a number ϵ such that $p^t u \geq \epsilon$ for all $u \in U$ and $p^t 0 < \epsilon$. That is, $p^t(x - y) \geq \epsilon$ for all $x \in S$ and for all $y \in T$. And hence $p^t x \geq \epsilon + p^t y$ for all $x \in S$ and for all $y \in T$. $\square$

Exercise 1.13. Prove or disprove: If S and T are two nonempty disjoint closed convex sets, then there exists a hyperplane that separates S and T.

Convex Cones, Polarity and Polyhedral Sets

Definition 1.9. A nonempty set C in R^n is called a cone with vertex zero if $x \in C$ implies $\lambda x \in C$ for all $\lambda \geq 0$. If, in addition, C is convex, then C is called a convex cone.

Draw the graphs of the following sets and check which of these are cones: (i) $S = \{\lambda(1,3) : \lambda \geq 0\}$, (ii) $S = \{\lambda(1,3) + \beta(2,1) : \lambda, \beta \geq 0\}$.

Exercise 1.14. Let S be a nonempty set in R^n. The polar cone of S, denoted by S^*, is defined as the set $\{p \in R^n : p^t x \leq 0 \text{ for all } x \in S\}$.

Lemma 1.2. Let S and T be nonempty sets in R^n. The following statements hold good:

1. S^* is a closed convex cone.

2. $S \subseteq S^{**}$, where S^{**} is the polar cone of S^*.

3. $S \subseteq T$ implies $T^* \subseteq S^*$.

Proof. 1. Let $x \in S^*$ and $\lambda \geq 0$. Let $y \in S$. By definition $x^t y \leq 0$ and hence $(\lambda x)^t y \leq 0$. Thus, $\lambda x \in S^*$ and S^* is a cone.

2. Let $x \in S$. To show that $x \in S^{**}$ we need to show that for any $y \in S^*$, $x^t y \leq 0$. Fix any $y \in S^*$. Since $y \in S^*$, for every $u \in S$, $y^t u \leq 0$. Since $x \in S$, $y^t x \leq 0$. Hence, $x \in S^{**}$ and $S \subseteq S^{**}$.

3. Let $y \in T^*$. To show that $y \in S^*$, we need to show that for any $x \in S$, $y^t x \leq 0$. Fix $x \in S$. Since $S \subseteq T$, $x \in T$. Since $y \in T^*$, $y^t x \leq 0$.$\square$

Theorem 1.9. Let C be a closed convex cone. Then $C = C^{**}$.

Proof. From the Lemma, $C \subseteq C^{**}$. Conversely, let $y \in C^{**}$. To the contrary assume that $y \notin C$. From Theorem 1.4, there exists a nonzero p such that $p^t y > p^t x$ for all $x \in C$. This implies $p^t x \leq 0$ for all $x \in C$ (for if $p^t x > 0$ for some $x \in C$, then we can choose a $\lambda > 0$ such that $\lambda p^x > p^t y$ which implies that in the inequality is flouted by λx which is in C as C is a cone). Hence, $p \in C^*$. Since $y \in C^{**}$, we must have $p^t y \leq 0$. Since $0 \in C$, we must also have $p^t y > 0$. From this contradiction, it follows that $C = C^{**}$. $\square$

Definition 1.10. Let A be an $m \times n$ real matrix and let $b \in R^n$. The set $S = \{x : Ax = b,\ x \geq 0\}$ is called a polyhedral set.

Definition 1.11. Let $S \subseteq R^n$ and $\bar{x} \in S$. Say that $\bar{x}$ is an extreme point of S if the following implication holds:

$$[x, y \in S,\ \lambda \in (0,1), \bar{x} = \lambda x + (1-\lambda)y] \Rightarrow x = y = \bar{x}.$$

Definition 1.12. Let $S \subseteq R^n$. A **nonzero** $d \in R^n$ is called a direction of S if $\forall\ x \in S$ and $\forall\ \lambda > 0$, $x + \lambda d \in S$. Two directions, d and f are said to be distinct if $d \neq \lambda f$ for any $\lambda > 0$. An extreme direction of S is a direction of S that cannot be written as a nonnegative linear combination of two distinct directions of S.

Theorem 1.10. Let A be an $m \times n$ real matrix of rank m and let $b \in R^m$. Let $S = \{x : Ax = b,\ x \geq 0\}$. Then, $\bar{x} \in S$ is an extreme point of S if, and only if, the columns of A corresponding to the the positive coordinates of $\bar{x}$ are linearly independent.

Theorem 1.11. Let A be an $m \times n$ real matrix of rank m and let $b \in R^m$. Let $S = \{x : Ax = b,\ x \geq 0\}$. Then, $d \in S$ is an extreme direction of S if, and only if, there exist a nonsingular submatrix B of A of order m, a column A_j of A not in B and a positive number μ such that $d_B = -\mu B^{-1} A_j,\ d_j = \mu$, and all other coordinates of d are zero.

Representation Theorem

Theorem 1.12. Let A be an $m \times n$ real matrix of rank m and let $b \in R^m$. Let $S = \{x : Ax = b,\ x \geq 0\}$. Let $x^1, x^2, \ldots, x^g$ be the extreme points and $d^1, d^2, \ldots, d^h$ be the extreme directions of the set S. Then,

$$S = \{\sum_{i=1}^{g} \lambda_i x^i + \sum_{j=1}^{h} \mu_j d^j : \lambda_i, \mu_j \geq 0\ \forall i, j, \text{ and } \sum_{i=1}^{g} \lambda_i = 1\}$$

Exercise 1.15. Show that S has finite number of extreme points and extreme directions.

Exercise 1.16. Show that (see Theorem 1.12) either

$$c^t x^i = \min\{c^t x : Ax = b, x \geq 0\} \text{ for some } i$$

$$\text{or} \quad c^t d^j < 0 \text{ for some } j.$$

Exercises

E1.17. Let S be a compact set in R^n. Show that $H(S)$ is closed. Is this result true if S is only closed and not bounded.

E1.18. Show that the system $Ax \leq 0$ and $c^t x > 0$ has a solution $x \in R^n$, where $A = \begin{bmatrix} 1 & -1 & -1 \\ 2 & 2 & 0 \end{bmatrix}$ and $c = (1, 0, 5)^t$.

E1.19. Let A be a $p \times n$ matrix and B be a $q \times n$ matrix. Show that exactly one of the following systems has a solution:

System 1. $Ax < 0,\ Bx = 0$ for some $x \in R^n$

System 2. $A^t u + B^t v = 0$ for some $(u, v),\ u \neq 0,\ v \geq 0$

E1.20. Let S and T be two nonempty convex sets in R^n. Show that there exists a hyperplane that separates S and T if, and only if, $\inf\{||x-y|| : x \in S,\ y \in T\} > 0$.

E1.21. Let S and T be two nonempty disjoint convex sets in R^n. Show that there exist nonzero vectors p and q such that

$$p^t x + q^t y \geq 0 \text{ for all } x \in S \text{ and all } y \in T.$$

E1.22. Let C and D be convex cones in R^n. Show that $C + D$ is a convex cone and that $C + D = H(C \cup D)$.

E1.23. Let C be a convex cone in R^n. Show that $C + C^* = R^n$. Is this representation unique?

Convex Functions

Convex function play an important role in optimization and development computational algorithms. In this chapter we shall look at convex function, their properties and some generalizations of convex functions. Concavity is like flip side of convexity. If a function f is convex, then $-f$ is concave. Almost all the results that we derive for convex functions can be stated for concave functions with necessary twists.Under convex functions we shall study

1. Basic definitions of convex functions
2. Subgradients of convex functions
3. Differentiable convex functions
4. Maxima and minima of convex functions
5. Generalizations of convex functions

Definition 2.1. Let S be a nonempty convex set in R^n. A function $f : S \to R$ is said to be convex on S if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \quad \text{for all } x, y \in S \text{ and all } \lambda \in (0,1)$$

Some examples of convex functions are

1. $f(x) = 3x + 4$
2. $f(x) = |x|$
3. $f(x) = x^2 - 2x$
4. $f(x) = x^t Ax$ where A is a positive semidefinite matrix of order n and $x \in R^n$.

The function $f(x) = -\sqrt{x}$ is a convex function on R_+.

Lemma 2.1. Let $f : S \to R$ be a convex function, where S is a convex subset of R^n. Then for any real number α, the level set f_α defined by $f_\alpha = \{x : f(x) \leq \alpha\}$ is a convex set.

Proof. Let $x, y \in f_\alpha$ and let $\lambda \in (0,1)$. Then, $f(x) \leq \alpha$ and $f(y) \leq \alpha$. By convexity of f,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$$

and hence $\lambda x + (1-\lambda)y \in f_\alpha$ $\Box$

Exercise 2.1. Show that $f : S \to R$ is a convex function if, and only if, for all $\lambda_i \in (0,1)$ with $\sum \lambda_i = 1$ and for all $x^i \in S$, $f(\sum \lambda_i x^i) \leq \sum \lambda_i f(x^i)$.

Continuity of Convex Functions

Theorem 2.1. Let S be a convex set in R^n with nonempty interior and let $f : S \to R$ be a convex function. Then f is continuous on the interior of S.

Proof. Let $\bar{x} \in int(S)$. Fix $\epsilon > 0$. Since $\bar{x} \in int(S)$, there exists a μ such that $||x - \bar{x}|| < \mu$ implies $x \in S$. Let $\theta = 1 + \max_i\{\max\{f(\bar{x} + \mu e^i) - f(\bar{x}),\ f(\bar{x} - \mu e^i) - f(\bar{x})\}\}$, where e^i is the i^{th} column of the identity matrix. Since f is convex and $\bar{x} = \frac{\bar{x} + \mu e^i + \bar{x} - \mu e^i}{2}$, either $f(\bar{x} + \mu e^i) - f(\bar{x}) \geq 0$ or $f(\bar{x} - \mu e^i) - f(\bar{x}) \geq 0$. Therefore, $\theta > 0$. Let $\delta = \min\{\frac{\mu}{n}, \frac{\epsilon\mu}{n\theta}\}$. Now, fix any x such that $||x - \bar{x}|| < \delta$. Will show that $||f(x) - f(\bar{x})|| < \epsilon$.

Let $\alpha_i = \frac{|x_i - \bar{x}_i|}{\mu}$. Then $\mu(\sum \alpha_i^2)^{\frac{1}{2}} = ||x - \bar{x}|| < \frac{\mu}{n}$ which in turn implies $\alpha_i \leq \frac{1}{n}$ for each i. Thus, for each i, $0 \leq n\alpha_i \leq 1$. Let z^i be the vector with its i^{th} coordinate as $\mu(x_i - \bar{x}_i)/|x_i - \bar{x}_i|$, and all other coordinates as zero. Then, $x = \bar{x} + \sum \alpha_i z^i$.

$$\begin{aligned}
f(x) &= f(\bar{x} + \sum \alpha_i z^i) \\
&= f(\frac{1}{n}\sum[\bar{x} + n\alpha_i z^i]) \\
&\leq \frac{1}{n}\sum f([\bar{x} + n\alpha_i z^i]) \\
&= \frac{1}{n}\sum f((1 - n\alpha_i)\bar{x} + n\alpha_i(\bar{x} + z^i)) \\
&\leq \frac{1}{n}\sum[(1 - n\alpha_i)f(\bar{x}) + n\alpha_i f(\bar{x} + z^i)] \\
&= f(\bar{x}) + \frac{1}{n}\sum n\alpha_i[f(\bar{x} + z^i) - f(\bar{x})] \qquad (1)
\end{aligned}$$

Note that $\alpha_i < \delta/\mu$ and hence $\sum \alpha_i < n\delta/\mu < \epsilon/\theta$. Therefore, rewriting (1), we get

$$f(x) - f(\bar{x}) \leq \frac{1}{n}\sum n\alpha_i\theta \leq \epsilon \qquad (2)$$

Next, let $y = 2\bar{x} - x$. Then, $||y - \bar{x}|| < \delta$ and hence

$$f(y) - f(\bar{x}) \leq \epsilon \tag{3}$$

Since $\bar{x} = \frac{1}{2}x + \frac{1}{2}y$, we have

$$f(\bar{x}) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) \tag{4}$$

Combining (3) and (4), we get $f(\bar{x}) - f(x) \leq \epsilon$. This completes the proof of the theorem. $\square$

Directional Derivative of Convex Functions

Definition 2.2. Let S be a nonempty set in R^n and let $f : S \rightarrow R$ be a function. Let $\bar{x} \in S$. A nonzero vector $d \in R^n$ is said to be a **feasible direction** of S at $\bar{x}$ if there exists a $\delta > 0$ such that $\bar{x} + \lambda d \in S$ for all $\lambda \in (0,\ \delta)$. Furthermore, for a feasible direction d of S at $\bar{x}$, f is said to have a **directional derivative** at $\bar{x}$ in the direction d if the following limit exists:

$$f(\bar{x}; d) = \lim_{\lambda \rightarrow 0^+} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda}$$

Note that we use the notation $f(\bar{x}; d)$ to denote the directional derivative of f at $\bar{x}$ in the direction d.

If the function f is convex and is defined globally (that is, $S = R^n$), then the directional derivative exists at all $x \in R^n$. However, when S is not whole of R^n, the directional derivative may not exist on ∂S.

Lemma 2.2. Let $f : R^n \rightarrow R$ be a convex function. Consider any point $\bar{x} \in R^n$ and a direction d. Then, $f(\bar{x}; d)$ exists.

Proof. Let $\lambda_2 > \lambda_1 > 0$. Then

$$\begin{aligned} f(\bar{x} + \lambda_1 d) &= f[\frac{\lambda_1}{\lambda_2}(\bar{x} + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})\bar{x}] \\ &\leq \frac{\lambda_1}{\lambda_2} f(\bar{x} + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(\bar{x}) \\ &\quad \text{(by convexity of } f\text{)} \end{aligned}$$

Rearranging the terms in the above inequality, we get

$$\frac{f(\bar{x} + \lambda_1 d) - f(\bar{x})}{\lambda_1} \leq \frac{f(\bar{x} + \lambda_2 d) - f(\bar{x})}{\lambda_2} \tag{5}$$

Let $g(\lambda) = \frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda}$. Then g is a nondecreasing function of λ over R_+.

Also by convexity of f, for any $\lambda >$, we have

$$
\begin{aligned}
f(\bar{x}) &= f[\frac{\lambda}{1+\lambda}(\bar{x}-d) + \frac{1}{1+\lambda}(\bar{x}+\lambda d)] \\
&\leq \frac{\lambda}{1+\lambda} f(\bar{x}-d) + \frac{1}{1+\lambda} f(\bar{x}+\lambda d) \qquad (6)
\end{aligned}
$$

Rearranging the terms in (6), we get

$$g(\lambda) = \frac{f(\bar{x}+\lambda) - f(\bar{x})}{\lambda} \geq f(\bar{x}) - f(\bar{x}-d).$$

Thus, $g(\lambda)$ is bounded below and hence the $\lim_{\lambda \to 0^+} g(\lambda)$ exists. □

Subgradients of Convex Functions

Definition 2.3. Let $f : S \to R$. The set $\{(x, f(x)) : x \in S\} \subseteq R^{n+1}$ is called the graph of the function. Furthermore, the sets $\{(x, y) : x \in S$ and $y \geq f(x)\}$ and $\{(x, y) : x \in S$ and $y \leq f(x)\}$ are called the epigraph and hypograph of f respectively.

We shall denote the epigraph of a function by $epi(f)$ and its hypograph by $hyp(f)$.

Theorem 2.2. Let S be a nonempty convex set of R^n and let $f : S \to R$ be a function. Then, f is convex if, and only if, $epi(f)$ is a convex set.

Proof. Assume f is convex. Let $(x, y), (u, v) \in epi(f)$ where $x, u \in S$. Let $\lambda \in (0, 1)$. Then, we have

$$f(x) \leq y \text{ and } f(u) \leq v.$$

Since f is convex,

$$f(\lambda x + (1-\lambda)u) \leq \lambda f(x) + (1-\lambda) f(u) \leq \lambda y + (1-\lambda)v.$$

Hence $\lambda(x, y) + (1-\lambda)(u, v) \in epi(f)$.

Conversely, assume that $epi(f)$ is convex. Let $x, u \in S$ and let $\lambda \in (0, 1)$. Let $y = f(x)$ and let $v = f(u)$. Then, (x, y) and (u, v) are in $epi(f)$. Since $epi(f)$ is convex, $\lambda(x, y) + (1-\lambda)(u, v) \in epi(f)$. Hence

$$\lambda y + (1-\lambda)v \geq f(\lambda x + (1-\lambda)u) \quad \text{or} \quad f(\lambda x + (1-\lambda)u) \leq \lambda f(x) + (1-\lambda) f(u).$$

It follows that f is convex. □

Definition 2.4. Let S be a nonempty convex set in R^n and let $f : S \to R$ be convex. Then a vector $\xi \in R^n$ is called a subgradient of f at a point $\bar{x} \in S$ if

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \quad \text{for all } x \in S. \tag{7}$$

Definition 2.5. Let S be a nonempty convex set in R^n and let $f : S \to R$. Say that f is concave on S if $-f$ is convex. If f is a concave function, then a vector $\xi \in R^n$ is called a subgradient of f at a point $\bar{x} \in S$ if

$$f(x) \leq f(\bar{x}) + \xi^t(x - \bar{x}) \quad \text{for all } x \in S. \tag{8}$$

Exercise 2.2. Analyze the convexity of the function $h(x) = \min\{f(x), g(x)\}$ where $f(x) = 4 - |x|$ and $g(x) = 4 - (x-2)^2$, $x \in \mathbf{R}$.

Theorem 2.3. Let S be a nonempty convex set of R^n and let $f : S \to R$ be a convex function. Then for every $\bar{x} \in int(S)$, f has a subgradient ξ at $\bar{x}$.

Proof. Since f is convex, $epi(f)$ is convex by Theorem 2.2. Note that the point $(\bar{x}, f(\bar{x}))$ is a point on the boundary of $epi(f)$. By Theorem 1.6, there exists a hyperplane that supports $epi(f)$ at $(\bar{x}, f(\bar{x}))$. That is, there will exist a nonzero (p, q) with $p \in R^n$ and $q \in \mathbf{R}$ such that

$$p^t\bar{x} + qf(\bar{x}) \geq p^t x + qy \quad \text{for all } x \in S \text{ and all } y \geq f(x). \tag{9}$$

We claim that $q \neq 0$. To the contrary, assume $q = 0$. Then from (9), we have $p^t\bar{x} \geq p^t x$ for all $x \in S$. As $\bar{x}$ is an interior point of S, $\bar{x} + \delta p \in S$ for all positive δ sufficiently small. This means $p^t\bar{x} \geq p^t(\bar{x} + \delta p)$ for a $\delta > 0$. This implies $p^t p \leq 0$ which in turn implies $p = 0$ leading to a contradiction that $(p, q) = 0$. Hence, it follows that $q \neq 0$. This clubbed with (9) implies that $q < 0$. Dividing both sides of (9) by q and letting $\xi = \frac{-1}{q}p$, we have

$$-\xi^t\bar{x} + f(\bar{x}) \leq -\xi^t x + y \quad \text{for all } x \in S \text{ and all } y \geq f(x).$$

Letting $y = f(x)$ in the above inequality we get

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \quad \text{for all } x \in S.$$

Thus, ξ is a subgradient of f at $\bar{x}$. $\square$

Definition 2.6. Let S be a nonempty convex set in R^n and let $f : S \to R$. Say that f is strictly convex on S if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \quad \text{for all } x, y \in S,\ x \neq y, \text{ and all } \lambda \in (0,1)$$

Corollary 2.1. Let S be a nonempty convex set of R^n and let $f : S \to R$ be a convex function. Then for every $\bar{x} \in int(S)$, there exists a ξ such that

$$f(x) > f(\bar{x}) + \xi^t(x - \bar{x}) \qquad \text{for all } x \in S \setminus \{\bar{x}\}.$$

Proof. By Theorem 2.2, there exists a ξ such that

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \tag{10}$$

for all $x \in S$. Assume, if possible, the equality holds for some $u \in S$ $f(u) = f(\bar{x}) + \xi^t(u - \bar{x})$. Let $\lambda \in (0, 1)$. By strict convexity of f, we have

$$\begin{aligned} f(\lambda u + (1-\lambda)\bar{x}) &< \lambda f(u) + (1-\lambda) f(\bar{x}) \\ &= \lambda f(\bar{x}) + \lambda \xi^t(u - \bar{x}) + (1-\lambda) f(\bar{x}) \\ &= f(\bar{x}) + \lambda \xi^t(u - \bar{x}) \end{aligned}$$

Taking $x = \lambda u + (1-\lambda)\bar{x}$ in (10), we have

$$f(\lambda u + (1-\lambda)\bar{x}) \geq f(\bar{x}) + \lambda \xi^t(u - \bar{x}) > f(\lambda u + (1-\lambda)\bar{x})$$

which is a contradiction. Corollary follows. $\square$

Theorem 2.4. Let S be a nonempty convex set of R^n and let $f : S \to R$. Suppose for every $x \in int(S)$, f has a subgradient ξ at x. That is, suppose for each $x \in int(S)$, there exists a ξ_x such that

$$f(u) \geq f(x) + \xi^t(u - x) \qquad \text{for all } u \in S.$$

Then f is convex on $int(S)$.

Proof. Let $x, y \in int(S)$ and let $\lambda \in (0, 1)$. Note that $\bar{x} = \lambda x + (1-\lambda)y$ is in $int(S)$. From the hypothesis, there exists a ξ such that

$$f(u) \geq f(\bar{x}) + \xi^t(u - \bar{x}) \qquad \text{for all } u \in S. \tag{11}$$

Note that $x - \bar{x} = (1-\lambda)(x - y)$ and $y - \bar{x} = -\lambda(x - y)$. Substituting x and y for u in (11) we get

$$f(x) \geq f(\bar{x}) + (1-\lambda)\xi^t(x - y) \tag{12}$$

and

$$f(y) \geq f(\bar{x}) - \lambda \xi^t(x - y) \tag{13}$$

Multiplying (12) by λ both sides and (13) by $(1-\lambda)$ both sides and adding the resulting inequalities, we get

$$\lambda f(x) + (1-\lambda)f(y) \geq f(\bar{x}) = f(\lambda x + (1-\lambda)y).$$

Hence, f is convex on $int(S)$. $\Box$

Exercise 2.3. Consider the function f defined on $S = \{(x,y) : 0 \leq x, y \leq 1\}$ as follows:

$$f(x,y) = \begin{cases} 0, & 0 \leq x \leq 1, 0 < y \leq 1; \\ \frac{1}{4} - (x - \frac{1}{2}), & 0 \leq x \leq 1, y = 0. \end{cases}$$

Is f a convex function? Does f have subgradient vectors at all interior points? If so, what are they?

Differentiable Convex Functions

Definition 2.7. Let S be a set in R^n with nonempty interior and let $f : S \to R$. Let $\bar{x} \in int(S)$. Say that f is differentiable at $\bar{x}$ if there exists a vector $\nabla f(\bar{x})$, called the gradient vector of f at $\bar{x}$, and there exists a function $\alpha : R^n \to \mathbf{R}$, such that

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) + ||x - \bar{x}||\alpha(\bar{x}, x - \bar{x})$$

where $\lim_{x \to \bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0$. If T is a open subset of S, then f is said to be differentiable on T if f is differentiable at each point in T.

Remark 2.1. When f is differentiable at $\bar{x}$, then the gradient vector is unique and is given by

$$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \frac{\partial f(\bar{x})}{\partial x_2}, \ldots\ldots, \frac{\partial f(\bar{x})}{\partial x_n} \right)^t$$

Theorem 2.5. Let S be a nonempty convex set of R^n and let $f : S \to R$ be convex. Suppose f is differentiable at $\bar{x} \in int(S)$. Then f has a unique subgradient at $\bar{x}$ and is equal to the gradient of f at $\bar{x}$.

Proof. Let ξ be a subgradient of f at $\bar{x}$ (exists by Theorem 2.3.) so that

$$f(x) \geq f(\bar{x}) + \xi^t(x - \bar{x}) \qquad \text{for all } x \in S. \tag{14}$$

Since f is differentiable at $\bar{x}$,

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) + ||x - \bar{x}||\alpha(\bar{x}, x - \bar{x}) \tag{15}$$

Subtracting (15) from (14), we get

$$(\xi - \nabla f(\bar{x}))^t(x - \bar{x}) - ||x - \bar{x}||\alpha(\bar{x}, x - \bar{x}) \leq 0 \tag{16}$$

Letting $x = \bar{x} + \delta(\xi - \nabla f(\bar{x}))$ in the above inequality (for sufficiently small $\delta > 0$), we get

$$\delta(\xi - \nabla f(\bar{x}))^t(\xi - \nabla f(\bar{x})) - \delta||\xi - \nabla f(\bar{x})||\alpha(\bar{x}, \delta(\xi - \nabla f(\bar{x}))) \leq 0 \tag{17}$$

Dividing both sides by δ and taking limit as $\delta \to 0^+$, we get

$$(\xi - \nabla f(\bar{x}))^t(\xi - \nabla f(\bar{x})) \leq 0$$

which implies $\xi = \nabla f(\bar{x})$. $\square$

Theorem 2.6. Let S be a nonempty open convex set of R^n and let $f : S \to R$ be differentiable on S. Then f is convex if, and only if, for every $\bar{x} \in S$

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) \quad \text{for all } x \in S.$$

Similarly, f os strictly convex if, and only if, for every $\bar{x} \in S$

$$f(x) > f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) \quad \text{for all } x \in S,\ x \neq \bar{x}.$$

Proof. Exercise.

Remark 2.2. Consider the optimization problem: Minimize $f(x)$, $x \in X$, where $X \subseteq R^n$. Note that the right hand sides of the inequalities of the above theorem provide lower bounds on f. Furthermore, the bounds are affine functions of x. This aspect is very useful in developing algorithms to solve the optimization problems.

Remark 2.3. Consider the system of nonlinear constraints defined by the set

$$X = \{x : g_i(x) \leq 0,\ i = 1, 2, .., m,\}$$

where g_is are differentiable convex functions on X. Instead of looking at the nonlinear system of constraints, we may first try and solve a linear system of constraints (linear program) by looking at the problem

$$Y = \{x : g_i(\bar{x}) + \nabla g_i(\bar{x})^t(x - \bar{x}) \leq 0,\ i = 1, 2, .., m.\}$$

Note that $X \subseteq Y$ as $x \in X$ implies $g_i(x) \leq 0$ and by the above theorem

$$g_i(\bar{x}) + \nabla g_i(\bar{x})^t (x - \bar{x}) \leq g_i(x) \leq 0,$$

which implies $x \in Y$. In other words, we first try to solve a linear program over a bigger set of X which is a polyhedral approximation of X and then try to push this bigger set towards X in successive iterations. Here Y is called a relaxation of X.

Theorem 2.7. Let S be a nonempty open convex set of R^n and let $f : S \to R$ be differentiable on S. Then f is convex if, and only if, for every $x, y \in S$, we have

$$[\nabla f(y) - \nabla f(x)]^t (y - x) \geq 0.$$

Similarly, f is strictly convex if, and only if, for every $x, y \in S,\ x \neq y$, we have

$$[\nabla f(y) - \nabla f(x)]^t (y - x) > 0.$$

Proof. Fix $x, y \in S$ and let $\bar{x} = \lambda x + (1 - \lambda) y$, where $\lambda \in (0, 1)$. From Theorem 2.6, f is convex if, and only if, for all $v \in S$,

$$f(u) \geq f(v) + \nabla f(v)^t (u - v) \quad \text{ for all } u \in S.$$

Substituting x and y for u and v in the above inequality, we

$$f(x) \geq f(y) + \nabla f(y)^t (x - y)$$

and

$$f(y) \geq f(x) + \nabla f(x)^t (y - x).$$

Combining these two inequalities, we can write

$$f(x) \geq f(x) + \nabla f(x)^t (y - x) + \nabla f(y)^t (x - y)$$

which is same as

$$[\nabla f(y) - \nabla f(x)]^t (y - x) \geq 0.$$

Therefore, convexity of f implies the above inequality.

Conversely, assume that for every $u, v \in S$, we have

$$[\nabla f(v) - \nabla f(u)]^t (v - u) \geq 0.$$

Fix $x, y \in S$. By Mean Value Theorem,

$$f(x) = f(y) + \nabla f(\bar{x})^t (x - y), \tag{18}$$

where $\bar{x} = \lambda x + (1 - \lambda) y$ for some $\lambda \in (0, 1)$.

From the hypothesis, we have

$$(1 - \lambda)[\nabla f(x) - \nabla f(\bar{x})]^t (x - y) \geq 0$$

Dividing by $(1 - \lambda)$ and substituting for $\nabla f(\bar{x})$ from (18), we get $f(y) \geq f(x) + \nabla f(x)^t (y - x)$. By an earlier theorem, it follows that f is convex. $\square$

Exercise 2.4. Prove the above theorem for strict convexity part.

Definition 2.8. Let S be a set in R^n with nonempty interior and let $f : S \to R$. Let $\bar{x} \in int(S)$. Say that f is twice differentiable at $\bar{x}$ if there exist a vector $\nabla f(\bar{x})$, a symmetric matrix $H(\bar{x})$, called the Hessian matrix, and a function $\alpha : R^n \to \mathbf{R}$, such that

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x}) + ||x - \bar{x}||^2 \alpha(\bar{x}, x - \bar{x})$$

for each $x \in S$ and $\lim_{x \to \bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0$.

Remark 2.4. When the function f is twice differentiable, the Hessian matrix is given by

$$H(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(\bar{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_n^2} \end{bmatrix}$$

Exercise 2.5. Find the Hessian matrix of $f(x, y) = 2x + 6y - 2x^2 - 3y^2 + 4xy$.

Exercise 2.6. Find the gradient and Hessian matrix of $f(x) = c^t x + x^t A x$ where A is an $n \times n$ matrix and $c, x \in R^n$.

Definition 2.9. A square matrix A of order n is said to be positive semidefinite (positive definite) if $x^t A x \geq 0$ for all x ($x^t A x > 0$ for all $x \neq 0$).

Theorem 2.8. Let S be a nonempty open convex set of R^n and let $f : S \to R$ be twice differentiable on S. Then f is convex if, and only if, the Hessian matrix of f is positive semidefinite at each point in S.

Proof. Assume that f is convex. Fix an $\bar{x} \in S$. Then for any $x \in R^n$, $\bar{x} + \lambda x \in S$ for all λ sufficiently small. We have

$$f(\bar{x} + \lambda x) \geq f(\bar{x}) + \lambda \nabla f(\bar{x})^t x \tag{19}$$

and by twice differentiability of f

$$f(\bar{x} + \lambda x) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t x + \frac{1}{2}\lambda^2 x^t H(\bar{x})x + \lambda^2 ||x||^2 \alpha(\bar{x}, \lambda x) \tag{20}$$

Combining (18) and (19) we get

$$\frac{1}{2}\lambda^2 x^t H(\bar{x})x + \lambda^2 ||x||^2 \alpha(\bar{x}, \lambda x) \geq 0$$

Dividing both sides by λ and taking limit as $\lambda \to 0$, we get $x^t H(\bar{x})x \geq 0$. As x was arbitrary, it follows that $H(\bar{x})$ is positive semidefinite.

Next, assume that $H(x)$ is positive semidefinite for all $x \in S$. Fix any $x, y \in S$. By mean value theorem, we have

$$f(x) = f(y) + \nabla f(y)^t (x - y) + \frac{1}{2}(x - y)^t H(\bar{x})(x - y),$$

where $\bar{x} = \lambda x + (1 - \lambda)y$ for some $\lambda \in (0, 1)$. Since $\bar{x} \in S$, $(x - y)^t H(\bar{x})(x - y) \geq 0$ and hence $f(x) \geq f(y) + \nabla f(y)^t (x - y)$ and by Theorem 2.6 it follows that f is convex. $\square$

Theorem 2.9. Let S be a nonempty open convex set of R^n and let $f : S \to R$ be twice differentiable on S. If the Hessian matrix of f is positive definite at each point in S, then f is strictly convex.

Exercise 2.7. Prove the above theorem.

Exercise 2.8. Let $f : R^n \to R$ be a function. For any $x, d \in R^n$, define the function $g_{(x,d)} : R \to \mathbf{R}$ by $g_{(x,d)}(\lambda) = f(x + \lambda d)$. Show

1. f is convex if, and only if, $g_{(x,d)}$ is convex for all $x \in R^n$ and for all nonzero $d \in R^n$.

2. f is strictly convex if, and only if, $g_{(x,d)}$ is strictly convex for all $x \in R^n$ and for all nonzero $d \in R^n$.

Minima and Maxima of Convex Functions

Let $f : R^n \to R$ be a function, where $S \subseteq R^n$. Consider the optimization problem: Minimize $f(x)$ subject $x \in S$.

Definition 2.10. Every $x \in S$ is called a feasible solution to the optimization problem. An $\bar{x} \in S$ is called a solution to the problem (also called called global optimal solution or simply optimal solution to the problem) if $f(x) \geq f(\bar{x})$ for all $x \in S$. An $\bar{x} \in S$ is said to be a **local optimal solution** (or **local minimum**) if there exists an $\epsilon > 0$ such that $f(x) \geq f(\bar{x})$ for all $x \in S$ with $||x - \bar{x}|| < \epsilon$. An $\bar{x} \in S$ is said to be a **strict local optimal solution** (or **strict local minimum**) if there exists an $\epsilon > 0$ such that $f(x) > f(\bar{x})$ for all $x \in S$ with $||x - \bar{x}|| < \epsilon$ and $x \neq \bar{x}$. An $\bar{x} \in S$ is said to be a **Strong optimal solution** (or **strong local minimum**) if there exists an $\epsilon > 0$ such that $\bar{x}$ is the only local optimal solution in $S \cap N_\epsilon(\bar{x})$, where $N_\epsilon(\bar{x})$ is the ϵ-neighbourhood of $\bar{x}$.

Exercise 2.9. Give an example to distinguish strict and strong local optimal solutions. Show that every strong local optimal solution is a strict optimal solution. Is the converse true?

Theorem 2.10. Let $f : S \to R$ be a convex function, where S is a nonempty convex subset of R^n. Suppose $\bar{x} \in S$ is a local optimal solution to the optimization problem: Minimize $f(x)$ subject $x \in S$.

1. Then, $\bar{x}$ is a global optimal solution.

2. If $\bar{x}$ is a strict local minimum, then $\bar{x}$ is unique global optimal solution.

3. If f is strictly convex, then $\bar{x}$ is unique global optimal solution.

Proof.
1. To the contrary, assume that $f(y) < f(\bar{x})$ for some $y \in S$. For $\lambda \in (0,1)$, $\lambda y + (1 - \lambda)\bar{x} \in S$, and by convexity of f,

$$\begin{aligned} f(\lambda y + (1-\lambda)\bar{x}) &\leq \lambda f(y) + (1-\lambda) f(\bar{x}) \\ &< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x}). \end{aligned}$$

For λ sufficiently close to 0, $\lambda y + (1-\lambda)\bar{x} \in S$ can be made arbitrarily close to $\bar{x}$ which will contradict local optimality of $\bar{x}$.

Exercise 2.10. Prove parts 2 and 3 of the above theorem.

Theorem 2.11. Let $f : R^n \to R$ be a convex function, where S is a nonempty convex subset of R^n. An $\bar{x} \in S$ is an optimal solution to the problem minimize $f(x)$ subject $x \in S$ if, and only if, f has a subgradient ξ at $\bar{x}$ such that $\xi^t(x-\bar{x}) \geq 0$ for all $x \in S$.

Proof. Suppose f has a subgradient ξ at $\bar{x}$ such that $\xi^t(x-\bar{x}) \geq 0$ for all $x \in S$. Then,

$$f(x) \geq f(\bar{x}) + \xi^t(x-\bar{x}) \text{ for all } x \in S.$$

Since $\xi^t(x-\bar{x}) \geq 0$ for all $x \in S$, we have $f(x) \geq f(\bar{x})$ for all $x \in S$ and hence $\bar{x}$ is a solution to the problem.

Conversely, assume that $\bar{x} \in S$ is an optimal solution to the problem. Define the sets

$$U = \{(x-\bar{x},\ y) : \ x \in R^n,\ y > f(x) - f(\bar{x})\}$$
$$V = \{(x-\bar{x},\ y) : \ x \in S,\ y \leq 0\}$$

Since S is convex, it follows that V is convex. Using convexity of f, it can be checked that U is convex (*Check this!*). Since $\bar{x}$ is an optimal solution to the problem, it follows that $U \cap V = \emptyset$. From Theorem 1.7, it follows that there exists a non-zero vector $(\beta,\ \mu)$ and a number α such that

$$\beta^t(x-\bar{x}) + \mu y \leq \alpha \text{ for all } x \in R^n \text{ and } y > f(x) - f(\bar{x}), \tag{21}$$

and

$$\beta^t(x-\bar{x}) + \mu y \geq \alpha \text{ for all } x \in S \text{ and } y \leq 0. \tag{22}$$

If $\mu > 0$, then (22) will be violated for large negative y. Hence $\mu \leq 0$. Letting $x = \bar{x}$ and $y = \epsilon > 0$ in (21), we get $\mu\epsilon \leq \alpha$. This implies, $\alpha \geq 0$. Taking $x = \bar{x}$ and $y = 0$ in (22), we get $\alpha \leq 0$. Hence $\alpha = 0$.

Suppose $\mu = 0$. Taking $x = \bar{x} + \beta$ in (21), we get $\beta^t\beta \leq 0$ which in turn implies $(\beta,\ \mu) = 0$, a contradiction. It follows that $\mu < 0$.

Dividing both sides of (22) by $-\mu$ and letting $\xi = -\frac{1}{\mu}\beta$, we get

$$\xi^t(x-\bar{x}) - y \geq 0 \quad \text{for all } x \in S \text{ and all } y \leq 0.$$

Taking $y = 0$ in the above inequality, we get

$$\xi^t(x - \bar{x}) \geq 0 \quad \text{for all } x \in S.$$

Since $\bar{x}$ is optimal $f(x) - f(\bar{x}) \geq 0$ for all $x \in S$. Dividing both sides of (21) by $-\mu$ and rearranging we get

$$\xi^t(x - \bar{x}) \leq y \quad \text{for all } x \in R^n \text{ and } y > f(x) - f(\bar{x}).$$

Taking $x \in S$ and taking limit as $y \to f(x) - f(\bar{x})$ we get,

$$\xi^t(x - \bar{x}) \leq f(x) - f(\bar{x}) \quad \text{for all } x \in S.$$

Thus, ξ is a subgradient of f at $\bar{x}$ with $\xi^t(x - \bar{x}) \geq 0$ for all $x \in S$. $\square$

Corollary 2.2. Let $f : R^n \to R$ be a function, where S is a nonempty open convex subset of R^n. An $\bar{x} \in S$ is an optimal solution to the problem minimize $f(x)$ subject $x \in S$ if, and only if, f has a zero subgradient at $\bar{x}$.

Proof. By the theorem, $\bar{x}$ is optimal if, and only if, $\xi^t(x - \bar{x}) \geq 0$ for all $x \in S$. Take $x = \bar{x} - \lambda\xi$ where $\lambda > 0$ such that $x \in S$ (S is open). This is possible, if, and only if, $\xi = 0$. $\square$

Corollary 2.3. Let $f : R^n \to R$ be a differentiable function, where S is a nonempty open convex subset of R^n. Then, an $\bar{x} \in S$ is an optimal solution to the problem of minimizing $f(x)$ subject $x \in S$ if, and only if, $\nabla f(\bar{x})^t(x - \bar{x}) \geq 0$ for all $x \in S$. In addition, if S is also open, then $\bar{x}$ is an optimal solution if, and only if, $\nabla f(\bar{x}) = 0$.

Proof. Exercise.

Method of Feasible Directions

Consider the problem: Minimize $f(x)$ subject $x \in S$. We shall assume that both S and f are convex. To find a solution to this problem, we start with an $\bar{x} \in S$. If $\bar{x}$ is not an optimal solution the problem, then we must have an $x \in S$ such that the function f must decrease in the direction $x - \bar{x}$ from $f(\bar{x})$ (otherwise, $\bar{x}$ will become a local optimal solution and by convexity $\bar{x}$ will also be a global optimal solution). Note that in this case, $x - \bar{x}$ will be a feasible direction of S at $\bar{x}$ (recall the Definition 2.2 of feasible direction). To obtain the direction, we solve the following optimization problem :

$$\text{Minimize } f[\bar{x} + \lambda(x - \bar{x})] \text{ subject to } x \in S. \tag{23}$$

Suppose f is a convex function over entire R^n and we are interested in the problem: Minimize $f(x)$ subject $x \in S$, where S is any arbitrary set, not necessarily convex. Again, let us start our search for an optimal solution from a point $\bar{x} \in S$. If $\bar{x}$ is an optimal solution to the problem, then any $y \in R^n$ with $f(y) < f(\bar{x})$ must not be in S. Let $y \in R^n$ be such that $f(y) < f(\bar{x})$. Then, if $\bar{x}$ is optimal to the problem, we must have

$$f(\bar{x}) > f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^t (y - \bar{x})$$

which implies that $\nabla f(\bar{x})^t (y - \bar{x}) < 0$. In other words, the hyperplane $H = \{u : \nabla f(\bar{x})^t (u - \bar{x}) = 0\}$ separates the set of all ys that are better than $\bar{x}$ (that is, $f(y) < f(\bar{x})$) from S. Thus, if $\bar{x}$ is an optimal solution, then we must have $\nabla f(\bar{x})^t (x - \bar{x}) \geq 0$ for all $x \in S$. Therefore, the problem reduces to

$$\text{Minimize } \nabla f(\bar{x})^t (x - \bar{x}) \text{ subject to } x \in S. \tag{24}$$

Note that (24) has a linear objective function, and if S is a polyhedral set, then the problem reduces to a Linear Programming problem.

Theorem 2.12. Consider the problem of minimizing $f(x)$ subject to $x \in S$, where f is a convex function and twice differentiable function, and S is a convex set, and suppose that there exists an optimal solution $\bar{x}$. Then, the set of alternative optimal solution to the problem is given by

$$V = \{x \in S : \ \nabla f(\bar{x})^t (x - \bar{x}) \leq 0, \ \text{ and } \ \nabla f(x) = \nabla f(\bar{x})\}$$

Proof. Let U be the set of all optimal solutions to the problem. Note that $U \neq \emptyset$ as $\bar{x} \in U$. Consider any $y \in V$. By the convexity of f and the definition of V, we have $y \in S$ and

$$f(\bar{x}) \geq f(y) + \nabla f(y)^t (\bar{x} - y) = f(y) + \nabla f(\bar{x})^t (\bar{x} - y) \geq f(y)$$

and hence we must have $y \in U$. Thus, $V \subseteq U$.

To prove the converse, let $y \in U$. Then, $y \in S$ and $f(y) = f(\bar{x})$. This means that $f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^t (y - \bar{x})$ or that $\nabla f(\bar{x})^t (y - \bar{x}) \leq 0$. Since $\bar{x}$ is optimal, we must have

$\nabla f(\bar{x})^t(y-\bar{x}) \geq 0$ and hence $\nabla f(\bar{x})^t(y-\bar{x}) = 0$. By interchanging the roles of y and $\bar{x}$, we must have $\nabla f(y)^t(y-\bar{x}) = 0$. Therefore,

$$[\nabla f(\bar{x}) - \nabla f(y)]^t(\bar{x}-y) = 0$$

Note that

$$\begin{aligned}
[\nabla f(\bar{x}) - \nabla f(y)] &= \nabla f[y+\lambda(\bar{x}-y)]_{\lambda=0}^{\lambda=1} \\
&= \int_{\lambda=0}^{\lambda=1} H[y+\lambda(\bar{x}-y)](\bar{x}-y)d\lambda = G(\bar{x}-y),
\end{aligned}$$

where $G = \int_{\lambda=0}^{\lambda=1} H[y+\lambda(\bar{x}-y)]d\lambda$ (the integral is performed each element-wise of the Hessian matrix. Observe that G is positive semidefinite. It follows that $(\bar{x}-y)^t G(\bar{x}-y) = 0$ which in turn implies $G(\bar{x}-y) = 0$ and hence $\nabla f(y) = \nabla f(\bar{x})$. $\Box$

Quasiconvex Functions

Definition 2.11. Let S be nonempty convex subset of R^n. A function from $f : S \to R$ is said to be quasiconvex if

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$$

$$\forall x, y \in S, \forall \lambda \in [0,1].$$

The function f is said be quasiconcave if $-f$ is quasiconvex.

Exercise 2.11. Show that f is quasiconvex if, and only if, its level sets are convex.

Theorem 2.13. Let S be nonempty open convex subset of R^n. Let $f : S \to R$ be a function differentiable on S. Then f is quasiconvex if, and only if, the following implication holds good:

$$\forall x, y \in S, \quad f(x) \leq f(y) \Rightarrow \nabla f(y)^t(x-y) \leq 0. \tag{25}$$

Proof. Assume f is quasiconvex. Fix $x, y \in S$. We may assume $f(x) \leq f(y)$. For any $\lambda \in (0,1)$, by differentiability of f,

$$f(\lambda x + (1-\lambda)y) - f(y) = \lambda \nabla f(y)^t(x-y) + \lambda \mid\mid x-y \mid\mid \alpha(y; \lambda(x-y)),$$

where $\alpha(y;\lambda(x-y)) \to 0$ as $\lambda \to 0$. Since f is quasiconvex, the LHS is nonpositive, and this implies

$$\lambda \nabla f(y)^t(x-y) + \lambda \mid\mid x-y \mid\mid \alpha(y;\lambda(x-y)) \leq 0.$$

Dividing both sides by λ and taking the limit as $\lambda \to 0$ we get

$$\nabla f(y)^t(x-y) \leq 0.$$

Conversely, assume that (25) holds. Need to show that f is quasiconvex. Take $x, y \in S$. We may assume $f(x) \leq f(y)$.Suppose there exists a $\lambda \in (0,1)$ such that

$$f(z) > f(y) \text{ where } z = \lambda x + (1-\lambda)y.$$

Since f is differentiable on S, it is continuous on S. This implies, there exists a $\delta \in (0,1)$ such that

$$f[\mu z + (1-\mu)y] > f(y) \ \forall \mu \in [\delta, 1] \quad (\text{ as } f(z) > f(y)),$$

$$f[\delta z + (1-\delta)y] < f(z)$$

The last inequality follows as $\delta z + (1-\delta)y$ is close to y for δ small, and $f(z) > f(y)$. By mean value theorem,

$$\nabla f(u)^t((1-\delta)(z-y)) = f(z) - f(\delta z + (1-\delta)y) > 0,$$

where $u = \mu z + (1-\mu)y$ for some $\mu \in (\delta, 1)$. This implies, as $z - y = \lambda(x-y)$, $\nabla f(u)^t(x-y) > 0$.
On the other hand, as $f(u) > f(y) \geq f(x)$, from (25), $\nabla f(u)^t(x-u) \leq 0$. As $x - u = (1-\lambda\mu)(x-y)$, the last inequality implies $\nabla f(u)^t(x-y) \leq 0$ which is a contradiction. It follows that f is quasiconvex. $\square$

Theorem 2.14. Let S be nonempty compact polyhedral set in R^n. Let $f : R^n \to \mathbf{R}$ be a quasiconvex and continuous function on S. Consider the problem of maximizing $f(x)$ subject to $x \in S$. There exists an optimal solution $\bar{x}$ to the problem which is an extreme point of S.

Proof. Since S is compact, it has no directions and every point of S is a convex combination of its extreme points. Let $x^1, x^2, \ldots, x^p$ be the extreme points of S. Let x^q be such that $f(x^q) = \max\{f(x^i) : 1 \leq i \leq p\}$. Given $x \in S$, we can write $x = \sum_{i=1}^{p} \lambda_i x^i$, a convex combination of extreme points of S. Note that as f is quasiconvex,

$$f(x) = f(\sum_{i=1}^{p} \lambda_i x^i) \leq \max\{f(x^i) : 1 \leq i \leq p\} = f(x^q).$$

Therefore, x^q is an optimal solution to the problem and proof is complete.

One of the sufficient conditions for a local optimal solution to be a global optimal solution is that the function f is *strictly quasiconvex.*

Definition 2.12. Let $f : S \to R$ be a function where S is a nonempty convex set in R^n. The function f is said to be strictly quasiconvex if for each $x, y \in S$ with $f(x) \neq f(y)$,

$$f(\lambda x + (1 - \lambda)y) < \max(f(x), f(y)) \; \forall \lambda \in (0, 1).$$

Theorem 2.15. Let S be nonempty convex set in R^n and let $f : R^n \to \mathbf{R}$ be a strictly quasiconvex function. Consider the problem of minimizing $f(x)$ subject to $x \in S$. If $\bar{x}$ is a local optimal solution to the problem, then it is also a global optimal solution to the problem.

Proof. Let $\bar{x}$ be a local optimal solution to the problem. Suppose $y \in S$ is such that $f(y) < f(\bar{x})$. For any $\lambda \in (0, 1)$, by strict convexity of f,

$$f[\bar{x} + \lambda(y - \bar{x})] < f(\bar{x}).$$

For λ sufficiently small, this will imply that $\bar{x} + \lambda(y - \bar{x})$ is locally better than $\bar{x}$, contradicting local optimality of $\bar{x}$. It follows that $\bar{x}$ is a global optimal solution to the problem. □

If a function is strictly convex, then it is also convex. However, a strictly quasiconvex function need not necessarily a quasiconvex function. Counterexample: Let $S = [-1, 1]$ and let $f(x) = 0 \; \forall \; x \neq 0$ and $f(0) = 1$.

Theorem 2.16. Let S be nonempty convex set in R^n and let $f : R^n \to \mathbf{R}$ be a strictly quasiconvex function. If f is continuous on S, then it is quasiconvex.

Proof. Let $x, y \in S$ be such that $f(x) = f(y)$. Suppose there exists a $\lambda \in (0,1)$ such that $f(\lambda x + (1-\lambda)y) > f(x)$. Let $z = \lambda x + (1-\lambda)y$. Since f is continuous, there exists a $\mu \in (0,1)$, such that $f(z) > f[\mu x + (1-\mu)z] > f(x) = f(y)$. Note that z is convex combination of $\mu x + (1-\mu)z$ and y. Since $f[\mu x + (1-\mu)z] > f(y)$, by strict quasiconvexity of f, $f(z) < f[\mu x + (1-\mu)z]$. From this contradiction it follows that we cannot find a $\lambda \in (0,1)$ such that $f(\lambda x + (1-\lambda)y) > f(x)$, and hence f is quasiconvex.

We have seen that a local minimum for a strictly quasiconvex function is also a global minimum. When can we say it is unique?

Definition 2.13. Let $f : S \to R$ be a function where S is a nonempty convex set in R^n. The function f is said to be strongly quasiconvex if for each $x, y \in S$ with $x \neq y$,

$$f(\lambda x + (1-\lambda)y) < \max(f(x), f(y)) \ \forall \lambda \in (0,1).$$

Note that

1. Every strictly convex function is strongly quasiconvex.
2. Every strongly quasiconvex function is strictly quasiconvex.
3. Every strongly quasiconvex function is quasiconvex.

Theorem 2.17. Let S be nonempty convex set in R^n and let $f : R^n \to \mathbf{R}$ be a strongly quasiconvex function. Consider the problem of minimizing $f(x)$ subject to $x \in S$. If $\bar{x}$ is a local optimal solution to the problem, then it is unique global optimal solution to the problem.

Proof. Exercise.

Definition 2.14. Let S be a nonempty open convex set in R^n and let $f : S \to R$ be a differentiable on S. The function f is said to be pseudoconvex on S if for each $x, y \in S$, the implication holds:

$$f(x) < f(y) \Rightarrow \nabla f(y)^t (x - y) < 0.$$

Theorem 2.18. Let S be nonempty open convex set in R^n and let $f : R^n \to \mathbf{R}$ be a differentiable pseudoconvex function. Then f is both strictly quasiconvex and quasiconvex.

Proof. We first show that f is strictly quasiconvex. To the contrary, assume $x, y \in S$ such that $f(x) \neq f(y)$ and $f(z) \geq \max\{f(x), f(y)\}$ where $z = \lambda x + (1-\lambda)y$ for some $\lambda \in (0,1)$. Assume, without loss of generality, $f(x) < f(y)$. Then, we have

$$f(z) \geq f(y) > f(x).$$

By pseudoconvexity of f, $\nabla f(z)^t(x - z) < 0$. Since $x - z = -(1-\lambda)(y-z)/\lambda$, this implies $\nabla f(z)^t(y - z) > 0$. This in turn implies, by pseudoconvexity of f, $f(y) \geq f(z)$ and hence $f(z) = f(y)$. Since $\nabla f(z)^t(y - z) > 0$, there exists a $u = \mu z + (1-\mu)y$ with $\mu \in (0,1)$ such that

$$f(u) > f(z) = f(y).$$

By pseudoconvexity of f, $\nabla f(u)^t(y - u) < 0$. Similarly, $\nabla f(u)^t(z - u) < 0$. Thus, we have

$$\nabla f(u)^t(y - u) < 0 \text{ and } \nabla f(u)^t(z - u) < 0.$$

But these two inequalities are contradicting each other as $y - u = \mu(u - z)/(1-\mu)$. It follows that f is strictly quasiconvex.
Since f is differentiable, it is continuous and hence quasiconvex.

Definition 2.15. Let S be a nonempty open convex set in R^n and let $f : S \to R$ be a differentiable on S. The function f is said to be strictly pseudoconvex on S if for each $x, y \in S$, the implication holds:

$$f(x) \leq f(y) \Rightarrow \nabla f(y)^t(x - y) < 0.$$

Theorem 2.19. Let S be nonempty open convex set in R^n and let $f : R^n \to \mathbf{R}$ be a differentiable strictly pseudoconvex function. Then f is strongly quasiconvex.

Proof. Exercise.

Exercises:

1. Suppose $f : R^n \to R$ is twice differentiable. If z is such that $\nabla f(z)$ vanishes, then show that $\lim_{\lambda \to 0} \frac{f(z+\lambda d) - f(z)}{\lambda^2}$ exists for any $d \in R^n$.
2. Show that every convex function is strictly quasiconvex as well as quasiconvex.
3. Show that every differentiable convex function is pseudoconvex.

4. Define various types of convexity at a point and examine which of the results developed so far hold good for functions having convexity (of different types) at a point.

5. Let $c, d \in R^n$ and let $\alpha, \beta \in \mathbf{R}$.

   Let $S = \{x : d^t x + \beta > 0\}$. Consider the function $f : S \to R$ defined by

$$f(x) = \frac{c^t x + \alpha}{d^t x + \beta}.$$

   Show that f is pseudoconvex.

3. NonLinear Programming
and
Necessary and Sufficient Conditions
for Optimality

- **Unconstrained Optimization**

  Minimize $f(x),\ x \in R^n$.

- **Constrained Optimization**

  Minimize $f(x)$ subject to $x \in S$ or

  With Inequality Constraints

  Minimize $f(x)$

  subject to

  $g_i(x) \leq 0,\ i = 1, 2, \ldots, m,$

  $x \in X \subseteq R^n$

  With Inequality and Equality Constraints

  Minimize $f(x)$

  subject to

  $g_i(x) \leq 0,\ i = 1, 2, \ldots, m,$

  $h_i(x) = 0,\ i = 1, 2, \ldots, l,$

  $x \in X \subseteq R^n$

Unconstrained Optimization

First Order Necessary Conditions for Optimality.

Theorem 3.1. Suppose $f : R^n \to \mathbf{R}$ is differentiable at $z \in R^n$. If there is a vector $d \in R^n$ such that $\nabla f(z)^t d < 0$, then there exists a $\delta > 0$ such that $f(z + \lambda d) < f(z)$ for each $\lambda \in (0, \delta)$, so that d is a *descent direction* of f at z.

Proof. Using differentiability of f at z, we can write

$$\frac{f(z + \lambda d) - f(z)}{\lambda} = \nabla f(z)^t d + \mid\mid d \mid\mid \alpha(z, \lambda d).$$

Since $\nabla f(z)^t d < 0$ and $\alpha(z, \lambda d) \to 0$ as $\lambda \to 0$, there exists a $\delta > 0$ such that the RHS of the above equation is negative for all $\lambda \in (0, \delta)$. The result follows. $\square$

Corollary 3.1. Suppose $f : R^n \to \mathbf{R}$ is differentiable at $z \in R^n$. If z is a local minimum, then $\nabla f(z) = 0$.

Proof. Since z is local minimum, for any d, $f(z + \lambda d) \geq f(z)$ $\forall \lambda$ sufficiently small which in turn implies $\nabla f(z)^t d \geq 0$ for all d. Take $d = -\nabla f(z)$. $\square$

Unconstrained Optimization

Second Order Necessary Conditions for Optimality.

Theorem 3.2. Suppose $f : R^n \to \mathbf{R}$ is twice differentiable at $z \in R^n$. If z is a local minimum, then $\nabla f(z) = 0$ and $H(z)$ is positive semidefinite.

Proof. Using differentiability of f at z and the hypothesis $\nabla f(z) = 0$, for any d we can write

$$\frac{f(z + \lambda d) - f(z)}{\lambda^2} = \frac{1}{2} d^t H(z) d + \| d \|^2 \alpha(z, \lambda d).$$

Since z is local minimum, $f(z + \lambda d) \geq f(z)$ $\forall \lambda > 0$ sufficiently small. Taking limit as $\lambda \to 0$, it follows that $d^t H(z) d \geq 0$ and hence the result follows. $\square$

Example 3.1. Minimize $f(x) = (x^2 - 1)^3, \ x \in \mathbf{R}$.
$\nabla f(x) = 6x(x^2 - 1)^2$; $\nabla f(x) = 0$ for $x = -1, 0, 1$. $H(x) = 24x^2(x^2 - 1) + 6(x^2 - 1)^2$ and $H(-1) = H(1) = 0$ and $H(0) = 6$. Verify that $z = 0$ is the local (global) minimum.

Unconstrained Optimization

Sufficient Conditions for Optimality

Theorem 3.3. Suppose $f : R^n \to \mathbf{R}$ is twice differentiable at $z \in R^n$. If $\nabla f(z) = 0$ and $H(z)$ is positive definite, then z is a local minimum.

Proof. If z is not local minimum, then there exists a sequence $x^k \to z$ such that $f(x^k) < f(z)$ for each k. Using the hypotheses, we can write

$$\frac{f(z + \lambda d_k) - f(z)}{\lambda^2} = \frac{1}{2} d_k^t H(z) d_k + \| d \|^2 \alpha(z, \lambda d_k),$$

where $d_k = (x^k - z)/ \parallel x_k - z \parallel$.
We may assume, without loss of generality, $d_k \to d$ for some $d \neq 0$. Taking limits as $\lambda \to 0$, we get $d^t H(z) d \leq 0$ which is a contradiction to positive definiteness of $H(z)$. $\square$
Reexamine Example 3.1.

Theorem 3.4. Suppose $f : R^n \to \mathbf{R}$ is pseudoconvex at z. Then z is global optimum if, and only if, $\nabla f(z) = 0$.

Optimization With Inequality Constraints

Consider the problem: Minimizing $f(x)$ subject to $x \in S$.
For any $z \in cl(S)$, the set of *feasible directions* of S at z is defined by

$$D(z) = \{d : d \neq 0, \text{ and } z + \lambda d \in S \ \forall \lambda \in (0, \delta) \text{for some } \delta > 0\}.$$

Similarly, define the set of *descent directions* of f at z by $F(z) = \{d : \nabla f(z)^t d < 0\}$. Note that $D(z)$ is a **cone** if $z \in S$. If z is local optimum, then $D(z) \cap F(z) = \emptyset$.

Optimization With Inequality Constraints

Theorem 3.5. Consider the problem of minimizing $f(x)$ subject to $x \in S$, where S is nonempty set in R^n and $f : R^n \to \mathbf{R}$ is differentiable at $z \in S$. If z is a local optimum solution to the problem, then $D(z) \cap F(z) = \emptyset$.

Proof. Suppose $d \in D(z) \cap F(z)$. This means we can find a $\lambda > 0$ arbitrarily small satisfying $f(z + \lambda d) < f(z)$ (because d is a direction of descent) and $z + \lambda d \in S$ (because d is a feasible direction). This contradicts the local optimality of z. It follows that $D(z) \cap F(z) = \emptyset$.

Consider the **Problem (PI)**: Minimize $f(x)$ subject to
$g_i(x) \leq 0,\ i = 1, 2, \ldots, m,\ x \in X \subseteq R^n$.
Here, each of f and g_is is a function from R^n to $\mathbf{R}$ and X is a nonempty open set in R^n.
So, the set of feasible solutions is given by

$$S = \{x \in X : g_i(x) \leq 0, i = 1, 2, \ldots, m\}.$$

Optimization With Inequality Constraints

When the NLP is specified as in the above problem, the necessary geometric condition for local optimality ($D(z) \cap F(z) = \emptyset$) can be reduced to an algebraic condition.

Theorem 3.6. Consider the problem PI stated above. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for i not in I are continuous at z. If z is a local optimal solution to the problem, then $F(z) \cap G(z) = \emptyset$, where $G(z) = \{d : \nabla g_i(z)^t d < 0,\ \forall i \in I\}$.

Proof. From an earlier result, $D(z) \cap F(z) = \emptyset$. Will show that $G(z) \subseteq D(z)$. Let $d \in G(z)$. As X is open, there exists a $\delta_1 > 0$ such that $z + \lambda d \in X\ \forall \lambda \in (0, \delta_1)$. For $i \notin I$, as $g_i(z) < 0$ and is continuous at z, there exists a $\delta_2 > 0$ such that $g_i(z + \lambda d) < 0\ \forall \lambda \in (0, \delta_2)$. For $i \in I$, as $\nabla g_i(z)^t d < 0$, d is a descent direction of g_i at z and hence $\exists$ a $\delta_3 > 0\ \ni g_i(z + \lambda d) < g_i(z) = 0\ \forall \lambda \in (0, \delta_3)$. From these inferences, we conclude that $d \in D(z)$ and hence the result follows.

Example 3.2.

$$
\begin{array}{ll}
\text{Minimize} & (x-3)^2 + (y-2)^2 \\
\text{subject to} & x^2 + y^2 \leq 5 \\
& x + y \leq 3 \\
& x, y \geq 0
\end{array}
$$

Analyze the optimality at the points $z = (\frac{9}{5}, \frac{6}{5})^t$ and $u = (2, 1)^t$.

$$\nabla f(z) = (\frac{-12}{5}, \frac{-8}{5})^t \text{ and } \nabla g_2(z) = (1, 1)^t.$$

Note that $F(z) \cap G(z) \neq \emptyset$ and hence z cannot be an optimal solution.

$$\nabla f(u) = (-2,\ -2)^t,\ \nabla g_1(u) = (4, 2)^t \text{ and } \nabla g_2(z) = (1, 1)^t.$$

Note that $F(u) \cap G(u) = \emptyset$ and hence u may be an optimal solution but this cannot be guaranteed from $F(u) \cap G(u) = \emptyset$ as it only a necessary condition.

Effect of the Form of Constraints

Utility of necessary conditions of the above theorem, i.e., $F(u) \cap G(u) = \emptyset$, may depend on how the constraints are expressed.

Example 3.3.

Minimize $(x-1)^2+(y-1)^2$

subject to $(x+y-1)^3 \leq 0$

$x, y \geq 0$

In this case, the necessary condition will hold good for each feasible (x, y) satisfying $x+y=1$. Now consider the same problem expressed as

Minimize $(x-1)^2+(y-1)^2$

subject to $x+y \leq 1$

$x, y \geq 0$

Verify that the necessary condition is satisfied only at the point $(\frac{1}{2}, \frac{1}{2})$.

Note that when $\nabla f(z)=0$ or $\nabla g_i(z)=0$ for $i \in I$, the necessary condition developed above is of no use.

Fritz John Conditions

Theorem 3.7. Consider the problem PI stated earlier. Suppose z is a feasible point to the problem. Let $I=\{i : g_i(z)=0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. If z is a local optimal solution to the problem, then there exist constants u_0 and u_i for $i \in I$ such that

$$\begin{aligned} u_0 \nabla f(z)+\sum_{i \in I} u_i \nabla g_i(z) &= 0 \\ u_0, u_i &\geq 0, \quad \text{for } i \in I \\ (u_0,\ u_I) &\neq 0 \end{aligned}$$

Furthermore, if g_is for $i \notin I$ are also differentiable at z, then there exist $u_0 \in \mathbf{R}$ and $u \in \mathbf{R}^m$ such that

$$u_0 \nabla f(z)+\sum_{i=1}^{m} u_i \nabla g_i(z) = 0 \tag{26}$$

$$u_i g_i(z) = 0, \text{ for } i=1,2,\ldots,m, \tag{27}$$

$$u_0, u_i \geq 0, \text{ for } i=1,2,\ldots,m, \tag{28}$$

$$(u_0,\ u^t) \neq 0 \tag{29}$$

Proof. Let $k = \mid I \mid$ and let A be the $n \times (k+1)$ matrix with its first column as $\nabla f(z)$ and its i^{th} column as $\nabla g_i(z),\ i \in I$. From the previous theorem, we know that

$F(z) \cap G(z) = \emptyset$. This is equivalent to saying there exists no d satisfying $\nabla f(z)^t d < 0$ and $\nabla g_i(z)^t d < 0$ for each $i \in I$. In other words, the system $A^t d < 0$ has no solution. By Gordan's theorem, there exists a nonzero nonnegative vector $p \in \mathbf{R}^{k+1}$ satisfying $Ap = 0$. Taking $u_0 = p_1,\ u_i = p_{i+1}$, the first assertion of the theorem follows. For the second assertion, take $u_i = 0$ for $i \notin I$.

The u_is in (26) in the statement of theorem are called the *Lagrangian multipliers*. The condition $u_i g_i(z) = 0,\ \ i = 1, 2, \ldots, m$, is called the *complementary slackness* condition.

Example 3.4.

Minimize	$(x-3)^2 + (y-2)^2$
subject to	$x^2 + y^2 \leq 5$
	$x + 2y \leq 4$
	$x, y \geq 0$

x	I	$\nabla f(x)$	$\nabla g_{i_1}(x)$	$\nabla g_{i_2}(z)$
$z = (2,1)^t$	$i_1 = 1,\ i_2 = 2$	$(-2,-2)^t$	$(4,2)^t$	$(1,2)^t$
$w = (0,0)^t$	$i_1 = 3,\ i_2 = 4$	$(-6,-4)^t$	$(-1,0)^t$	$(0,-1)^t$

Example 3.5.

Minimize	$-x$
subject to	$y - (1-x)^3 \leq 0$
	$y \geq 0$

Note that $z = (1,0)^t$ is the optimal solution to the problem (draw the feasible region and check this) and the Fritz John conditions hold good at this point. Here $I = \{1,2\}$, $\nabla f(z) = (-1,0)^t$, $\nabla g_1(z) = (0,1)^t$ and $\nabla g_2(z) = (0,-1)^t$. For the Fritz John condition we must have

$$u_0 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which holds good only if $u_0 = 0$.

Example 3.6.

Minimize	$-x$
subject to	$x + y \leq 0$
	$y \geq 0$

Note that Fritz John condition holds good at $z = (1, 0)^t$ with $u_0 = u_1 = u_2 = \alpha$ for any real $\alpha > 0$.
$\nabla f(z) = (-1, 0)^t$ $\nabla g_1(z) = (1, 1)^t$, $\nabla g_2(z) = (0, -1)^t$.

Kuhn-Tucker Necessary Conditions

Note that in examples 3.4 and 3.6, u_0 is positive. But in example 3.5, $u_0 = 0$. In example 3.5, the $\nabla g_i(z)$s for $i \in I$ are linearly dependent, but not in the other two examples. Note that when $u_0 = 0$ in Fritz John condition, the condition only talks about the constraints. With an additional assumption, the Fritz John condition can be improved. This is due to Kuhn and Tucker.

Theorem 3.8. Consider the problem PI stated earlier. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. Also assume that g_is for $i \in I$ are linearly independent. If z is a local optimal solution to the problem, then there exist constants u_i for $i \in I$such that

$$\begin{aligned} \nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) &= 0 \\ u_i &\geq 0, \quad \text{for } i \in I. \end{aligned}$$

Furthermore, if g_is for $i \notin I$ are also differentiable at z, then there exist $u \in \mathbf{R}^m$ such that

$$\nabla f(z) + \sum_{i=1}^{m} u_i \nabla g_i(z) = 0 \tag{30}$$
$$u_i g_i(z) = 0, \text{ for } i = 1, 2, \ldots, m, \tag{31}$$
$$u_i \geq 0, \text{ for } i = 1, 2, \ldots, m. \tag{32}$$
$$\tag{33}$$

Proof. Get u_0 and u_is as in the previous theorem. Note that $u_0 > 0$, as $\nabla g_i(z)$s for $i \in I$ would become linearly dependent otherwise. Since $u_0 > 0$, we can as well assume that it is equal to one without loss of generality. The second assertion of the theorem can be established as in the previous theorem.

Note that a geometric interpretation of the Kuhn-Tucker conditions is that if z is a local optimum, then the gradient vector of the objective function at z with its sign reversed is contained in the cone generated by the gradient vectors of the binding constraints (follows from (30) above).

Kuhn-Tucker Sufficient Conditions

Theorem 3.9. Consider the problem PI stated earlier. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. Further, assume that f is pseudoconvex at z and g_i is differentiable quasiconvex at z for each $i \in I$. If the Kuhn-Tucker conditions hold good at z, that is, there exist $u_i \in \mathbf{R}$ for each $i \in I$ such that $\nabla f(z) + \sum_{i=1}^{m} u_i \nabla g_i(z) = 0$, then z is a global optimal solution to the problem.

Proof. Let x be any feasible solution to PI. Then, for $i \in I$, $g_i(x) \leq 0 = g_i(z)$. By quasiconvexity of g_i at z,

$$g_i(z + \lambda(x - z)) = g_i(\lambda x + (1 - \lambda)z) \leq \max\{g_i(x), g_i(z)\} = g_i(z)$$

for all $\lambda \in (0, 1)$. Thus g_i does not increase in the direction $x - z$ and hence we must have $\nabla g_i(z)^t(x - z) \leq 0$. This implies $\sum_{i \in I} u_i \nabla g_i(z)^t(x - z) \leq 0$. Since $\nabla f(z) + \sum_{i=1}^{m} u_i \nabla g_i(z) = 0$, it follows that $\nabla f(z)^t(x - z) \geq 0$. Since f is pseudoconvex, $f(x) \geq f(z)$. It follows that z is global minimum.

Consider the **Problem (PIE)**:
Minimize $f(x)$
subject to

$g_i(x) \leq 0, \ i = 1, 2, \ldots, m,$

$h_i(x) = 0, \ i = 1, 2, \ldots, l, \qquad x \in X \subseteq R^n.$

Here, each of f, g_is and h_is is a function from R^n to $\mathbf{R}$ and X is a nonempty open set in R^n.
So, the set of feasible solutions is given by

$$S = \{x \in X : g_i(x) \leq 0, i = 1, \ldots, m; h_i(x) = 0, i = 1, \ldots, l\}.$$

Treating Equalities as Inequalities

Consider the problem: Minimize $f(x)$ subject to

$g(x) = 0,\ x \in X,\ X$ is a nonempty subset in R^n.

Letting $g_1(x) = g(x)$ and $g_2(x) = -g(x)$, the above problem can be stated as Minimize $f(x)$ subject to

$g_i(x) \leq 0,\ i = 1, 2,\ x \in X.$

Note that, $G(z) = \emptyset$ is true and hence the optimality conditions developed above are of no use.

Optimality Conditions for PIE

Theorem 3.10. Consider the problem PIE stated above. Suppose z is a local optimal solution to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. Further, assume that h_i is continuously differentiable at z for $i = 1, 2, \ldots, l$. If $\nabla h_i(z)$s, $i = 1, 2, \ldots, l$, are linearly independent, then $F(z) \cap G(z) \cap H(z) = \emptyset$, where

$F(z) = \{d : \nabla f(z)^t d < 0\}.$

$G(z) = \{d : \nabla g_i(z)^t d < 0,\ \forall i \in I\}.$

$H(z) = \{d : \nabla h_i(z)^t d = 0,\ \text{for} = 1, 2, \ldots, l\}.$

Fritz John Necessary Conditions

Theorem 3.11. Consider the problem PIE. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. Further, assume that h_i is continuously differentiable at z for $i = 1, 2, \ldots, l$. If z is a local optimal solution to the problem, then there exist constants u_0, u_i for $i \in I$ and $v_i,\ i = 1, 2, \ldots l$ such that

$$\begin{aligned} u_0 \nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) &= 0 \\ u_0, u_i &\geq 0, \quad \text{for } i \in I \\ (u_0,\ u_I, v) &\neq 0, \end{aligned}$$

where u_I is the vector of u_is corresponding to I and $v = (v_1, \ldots, v_l)^t$.

Furthermore, if g_is for $i \notin I$ are also differentiable at z, then there exist $u_0 \in \mathbf{R}$ and $u \in \mathbf{R}^m$ such that

$$u_0 \nabla f(z) + \sum_{i=1}^{m} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) \quad = \quad 0 \tag{34}$$

$$u_i g_i(z) = 0, \text{ for } i = 1, \ldots, m, \tag{35}$$

$$u_0, u_i \geq 0, \text{ for } i = 1, \ldots, m, \tag{36}$$

$$(u_0,\ u^t, v^t) \neq 0, \tag{37}$$

Proof. If $\nabla h_i(z), i = 1, \ldots, l$ are lineraly dependent, then there exist $v_i,\ i = 1, \ldots, l$, not all of them equal to zero, such that $\sum_{i=1}^{l} v_i \nabla h_i(z) = 0$. Taking u_0 and u_is to be zero, we see that z satisfies the necessary conditions.

Suppose $\nabla h_i(z), i = 1, \ldots, l$ are lineraly independent. Let A be the matrix whose first column is $\nabla f(z)$ and the remaining columns being $\nabla g_i(z),\ i \in I$. Let B be the matrix whose i^{th} column is $\nabla h(z),\ i = 1, \ldots, l$. Then from the previous theorem, there is no d which satisfies

$$A^t d < 0 \text{ and } B^t d = 0.$$

Define the sets $S = \{(p, q) : p = A^t d, q = B^t d, d \in R^n\}$ and $T = \{(p, q) : p < 0,\ q = 0\}$. Note that S and T are disjoint convex sets. Therefore, there exists a vector $(u_0, u_I^t, v^t) \neq 0$ such that

$$(u_0, u_I^t) A^t d + v^t B^t d \geq (u_0, u_I^t) p + v^t q \forall d \in R^n,\ \forall (p, q) \in cl(T).$$

Since $(p, 0) \in Cl(T)$ for arbitrarily large negative, it follows that $(u_0, u_I^t) \geq 0$. Since $(0, 0) \in cl(T)$,
$(u_0, u_I^t) A^t d + v^t B^t d \geq 0$ for all $d \in R^n$.
This implies $(u_0, u_I^t) A^t + v^t B^t = 0$. From this the theorem follows.

Remark 3.1. Note that the Lagrangian multipliers associated with h_is are unrestricted in sign.

Exercise. Write the Fritz John's conditions in the vector notation.

Example 3.7.

$$
\begin{array}{ll}
\text{Minimize} & (x-3)^2+(y-2)^2 \\
\text{subject to} & x^2+y^2 \le 5 \\
 & x+2y=4 \\
 & x, y \ge 0
\end{array}
$$

Analyze at $z(2,1)^t$.

Example 3.8.

$$
\begin{array}{ll}
\text{Minimize} & -x \\
\text{subject to} & y-(1-x)^3=0 \\
 & y \ge 0
\end{array}
$$

Analyze at $z(1,0)^t$.

Kuhn-Tucker Necessary Conditions

Theorem 3.12. Consider the problem PIE. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Assume f and g_is for $i \in I$ are differentiable at z and that g_is for $i \notin I$ are continuous at z. Further, assume that h_i is continuously differentiable at z for $i = 1, 2, \ldots, l$. Also assume that ∇g_is for $i \in I$ and $\nabla h_i(z)$s, $i = 1, \ldots, l$, are linearly independent.

If z is a local optimal solution to the problem, then there exist constants u_i for $i \in I$ and $v_i,\ i = 1, 2, \ldots l$ such that

$$\begin{aligned} \nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) &= 0 \\ u_i &\geq 0, \quad \text{for } i \in I, \end{aligned}$$

where u_I is the vector of u_is corresponding to I and $v = (v_1, \ldots, v_l)^t$.

Furthermore, if g_is for $i \notin I$ are also differentiable at z, then there exist $u \in \mathbf{R}^m$ and $v \in \mathbf{R}^l$ such that

$$\nabla f(z) + \sum_{i=1}^{m} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) = 0 \tag{38}$$

$$u_i g_i(z) = 0, \text{ for } i = 1, \ldots, m, \tag{39}$$

$$u_i \geq 0, \text{ for } i = 1, \ldots, m, \tag{40}$$

Example 3.9.

Minimize	$-x$
subject to	$x + y \leq 0$
	$y \geq 0$

Kuhn-Tucker Sufficient Conditions

Theorem 3.13. Consider the problem PIE. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Suppose that the Kuhn-Tucker conditions hold good at z, i.e., there exist scalers $u_i,\ i \in I$ and $v_i,\ i = 1, \ldots, l$, such that

$$\nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) = 0.$$

Let $J = \{i : v_i > 0\}$ and $K = \{i : v_i < 0\}$. Assume that f is pseudoconvex at z, g_is are quasiconvex at z for $i \in I$, h_is are quasiconvex at z for $i \in J$ and that h_is are quasiconcave at z for $i \in K$. Then z is a global optimal solution to the problem.

Kuhn-Tucker Sufficient Conditions

Theorem 3.14. Consider the problem PIE. Suppose z is a feasible point to the problem. Let $I = \{i : g_i(z) = 0\}$. Suppose that the Kuhn-Tucker conditions hold good at z, i.e., there exist scalers $u_i,\ i \in I$ and $v_i,\ i = 1, \ldots, l$, such that

$$\nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) = 0.$$

Let $J = \{i : v_i > 0\}$ and $K = \{i : v_i < 0\}$. Assume that f is pseudoconvex at z, g_is are quasiconvex at z for $i \in I$, h_is are quasiconvex at z for $i \in J$ and that h_is are quasiconcave at z for $i \in K$. Then z is a global optimal solution to the problem.

Proof. Let x be any feasible solution to PIE. Then, for $i \in I$, $g_i(x) \leq 0 = g_i(z)$. By quasiconvexity of g_i at z,

$$g_i(z + \lambda(x - z)) = g_i(\lambda x + (1 - \lambda)z) \leq \max\{g_i(x), g_i(z)\} = g_i(z)$$

for all $\lambda \in (0, 1)$. Thus g_i does not increase in the direction $x - z$ and hence we must have $\nabla g_i(z)^t(x - z) \leq 0$. This implies $\sum_{i \in I} u_i \nabla g_i(z)^t(x - z) \leq 0$.
Similarly, using quasiconvexity of h_is for $i \in J$ and quasiconcavity of h_is for $i \in K$, we can show that

$$\nabla h_i(z)^t(x - z) \leq 0 \text{ for } i \in J \text{ and}$$

$$\nabla h_i(z)^t(x - z) \geq 0 \text{ for } i \in K.$$

From the above inequalities, we can conclude

$$[\sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z)]^t(x - z) \leq 0.$$

Since $\nabla f(z) + \sum_{i \in I} u_i \nabla g_i(z) + \sum_{i=1}^{l} v_i \nabla h_i(z) = 0$ it follows that $\nabla f(z)^t(x - z) \geq 0$. Since f is pseudoconvex, $f(x) \geq f(z)$. It follows that z is global minimum.

4. Duality in NLP

We shall refer to PIE as the **Primal Problem.** We shall write this problem in the vector notation as

Minimize $f(x)$

subject to

$g(x) \leq 0,$

$h(x) = 0, \qquad x \in X \subseteq R^n.$

Here $g(x) = (g_1(x), g_2(x), \ldots, g_m(x))^t$ and

$h(x) = (h_1(x), h_2(x), \ldots, h_l(x))^t$

A number of problems closely associated with primal problem, called the dual problems, have been proposed in the literature. Lagrangian Dual problem is one of these problems which has played a significant role in the development of algorithms for

- large-scale linear programming problems,
- convex and nonconvex nonlinerar problems,
- discrete optimization problems.

The Lagrangian Dual Problem of PIE

Maximize $\theta(u, v)$

subject to

$u \geq 0$, where

$$\theta(u, v) = \inf\{f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x) : x \in X\}.$$

In the vector notation, the Lagrangian dual is written as Maximize $\theta(u, v)$

subject to $u \geq 0$,

where $\theta(u, v) = \inf\{f(x) + u^t g(x) + v^t h(x) : x \in X\}$.

Since the dual maximizes the infimum, the dual is sometimes called the max-min problem.

Note that

- the lagrangian dual objective function $\theta(u, v)$ incorporates constraint functions of the primal, the objective function of the primal, and the lagrangian multipliers of the primal encountered in the optimality conditions,
- the lagrangian multipliers associated with '$\leq$' constraints ($g(x) \leq 0$), namely the u_is are nonnegative and those associated with the '$=$' constraints ($h(x) = 0$), namely v_is are unrestricted in sign,
- the lagrangian dual objective function $\theta(u, v)$ may be $-\infty$ for a fixed vector (u, v), because it is the infimum of a functional expression over a set X,
- the lagrangian dual of a PIE is generally not unique as it depends on which constraints we treat as g_is, which constraints as h_is and which constraints as X,
- the choice of a lagrangian dual would affect the solution process using the dual approach to solve the primal.

Geometric Interpretation of Lagrangian Dual

Consider the problem
Minimize $f(x)$
subject to
$g(x) \leq 0,$
$x \in X \subseteq R^n.$
Here, both f and g are functions from R^n to $\mathbf{R}$. There is only one inequality constraint and no equality constraints.
For each $x \in X$, the two-tuple $(g(x),\ f(x))$ can be plotted on the two dimensional plane.
Let $G = \{(z_1,\ z_2) : z_1 = g(x),\ z_2 = f(x),\ x \in X\}$. A solution to the primal problem is that x which corresponds to $(z_1,\ z_2)$ in G such that $z_1 \leq 0$ and z_2 is minimum.
The lagrangian dual objective function for this problem is given by $\theta(u) = \inf\{f(x) + ug(x) : x \in X\}$, where u is nonnegative. That is, $\theta(u) = \inf\{z_2 + uz_1) : (z_1,\ z_2) \in G\}$, where u is nonnegative.

Note that for each $u \geq 0$ fixed, the dual objective value is the intercept of the line z_2+uz_1 supporting G (from below) on the z_2 axis. Therefore, the dual problem is equivalent to finding the slope of the supporting hyperplane of G such that the intercept on the z_2 axis is maximal. Note that at $(\bar{z}_1, \bar{z}_2)$ the dual objective function attains its maximum with $\bar{u}$ as the dual optimal solution. Also, in this case, the optimum dual objective value coincides with the optimum primal objective value.

Example 4.1.

Minimize $\quad f(x,y) = x^2 + y^2$

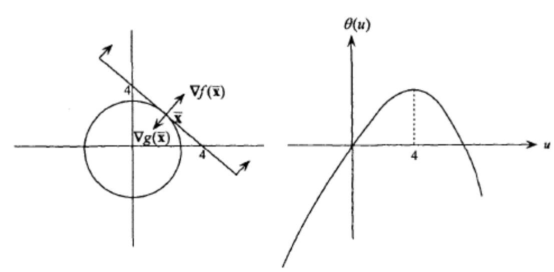
subject to $\quad g(x,y) = -x - y + 4 \leq 0$

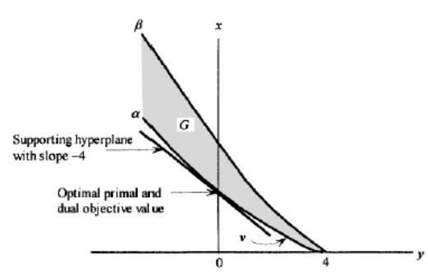
$\qquad\qquad\quad x,\ y \geq 0$

Verify that $(2,2)^t$ is the optimum solution to this problem with optimal objective value equal to 8.

Taking $X = \{(x,y) : x \geq 0,\ y \geq 0\}$, the dual function is given by

$$\begin{aligned}
\theta(u) &= \inf\{x^2 + y^2 + u(-x-y+4) : x \geq 0,\ y \geq 0\} \\
&= \inf\{x^2 - ux : x \geq 0\} + \inf\{y^2 - uy : y \geq 0\} + 4u \\
&= \begin{cases} -\frac{1}{2}u^2 + 4u, & \text{for } u \geq 0\ ; \\ 4u, & \text{for } u < 0. \end{cases}
\end{aligned}$$





Duality Theorems and Saddle Point Optimality

Theorem 4.1 (The Weak Duality Theorem). Let x be a feasible solution to PIE an let (u, v) be a solution to its Lagrangian dual. The $f(x) \geq \theta(u, v)$.

Corollary 4.1. $\inf\{f(x) : x \in X,\ g(x) \leq 0,\ h(x) = 0\} \geq \sup\{\theta(u, v) : u \geq 0\}$.

Corollary 4.2. If x is a feasible solution to PIE and (u, v) is a solution to its Lagrangian dual such that $f(x) \leq \theta(u, v)$, then x is optimal for PIE and (u, v) is optimal for the dual.

Corollary 4.3. If $\inf\{f(x) : x \in X,\ g(x) \leq 0,\ h(x) = 0\} = -\infty$, then $\theta(u, v) = -\infty$ for each $u \geq 0$.

Corollary 4.4. If $\sup\{\theta(u, v) : u \geq 0\} = \infty$, then PIE has no feasible solution.

Duality Gap

Remark: Note that in PIE, it was assumed that the set X was a nonempty open set. However, for the dual formulation, the openness of X is not required. In fact, X may even be a dicrete/finite set. Check that the weak duality theorem and its corollaries hold good for any nonempty set X. Henceforth, we shall refer to PIE without the openness assumption of X as the **primal problem**.

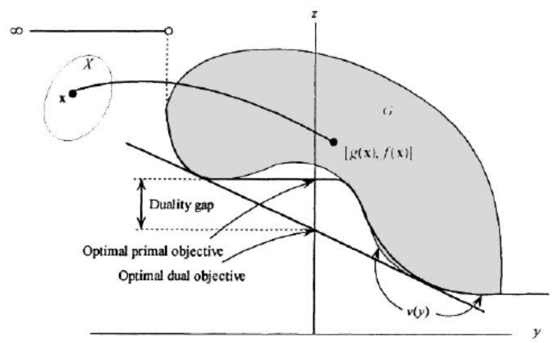
The dual optimal objective value may be a strictly less than the primal optimal objective value. In this case we say that there is *duality gap.* Analyze the following example.

Example 4.2.

Minimize $\quad f(x,y) = -2x + y$

subject to $\quad h(x,y) = x + y - 3 = 0$

$(x,\ y) \in X = \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$.



The *Strong Duality Theorem* asserts that, under some convexity assumptions and a constraint qualification, the primal optimal objective value is equal to the dual optimal objective value.

Theorem 4.2 (The Strong Duality Theorem). Let X be a nonempty convex set in R^n, let $f : R^n \to \mathbf{R}$ and $g : R^n \to \mathbf{R}^m$ be convex, and let $h : R^n \to \mathbf{R}^l$ be affine, that is, h is of the form $h(x) = Ax - b$. Suppose that the following constraint qualification holds true. There exists a $z \in X$ such that $g(z) < 0$ and $h(z) = 0$, and $0 \in int(h(X))$, where $h(X) = \{h(x) : x \in X\}$. Then

$$\inf\{f(x) : x \in X,\ g(x) \leq 0,\ h(x) = 0\} = \sup\{\theta(u,v) : u \geq 0\}$$

Furthermore, if the infimum is finite, then the $\sup\{\theta(u,v) : u \geq 0\}$ is attained at $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$. If the infimum is attained at $\bar{x}$, then $\bar{u}^t g(\bar{x}) = 0$.

Lemma: Let X be a nonempty convex set in R^n, let $\alpha : R^n \to \mathbf{R}$ and $g : R^n \to \mathbf{R}^m$ be convex, and let $h : R^n \to \mathbf{R}^l$ be affine. Consider the two systems:

System 1. $\alpha(x) < 0,\ g(x) \leq 0,\ h(x) = 0$ for some $x \in X$

System 2. $u_0\alpha(x) + u^t g(x) + v^t h(x) \geq 0\ \forall x \in X$

$(u_0, u) \geq 0,\ (u_0, u, v) \neq 0$

If System 1 has no solution, then System 2 has a solution. The converse is true if $u_0 > 0$.

Proof. Suppose that System 1 has no solution. Consider the set

$$\Lambda = \{(p, q, r) : p > \alpha(x),\ q \geq g(x),\ r = h(x) \text{ for some } x \in X\}$$

Since α and g are convex and h is affine, Λ is convex. Since System 1 has no solution, the vector $(0, 0, 0) \in \mathbf{R}^{1+m+l}$ does not belong to Λ. By a separation theorem, there exists a non-zero vector $(u_0, u, v) \in \mathbf{R}^{1+m+l}$ such that

$$u_0 p + u^t q + v^t r \geq 0\ \forall\ (p, q, r) \in cl(\Lambda) \tag{41}$$

Fix any $x \in X$. Note that $(\alpha(x), g(x), h(x)) \in cl(\Lambda)$ and $(p, q, h(x)) \in cl(\Lambda)$ for all $(p, q) > (\alpha(x), g(x))$.

From this and (16), it follows that $(u_0, u) \geq 0$. It follows that (u_0, u, v) is a solution to System 2.

To prove the converse, assume System 2 has a solution (u_0, u, v) with $u_0 > 0$.

Let $x \in X$ be such that $g(x) \leq 0$ and $h(x) = 0$. Since (u_0, u, v) solves System 2, we have

$$u_0\alpha(x) + u^t g(x) + v^t h(x) \geq 0.$$

Since $g(x) \leq 0$, $h(x) = 0$ and $u \geq 0$, it follows $u_0\alpha(x) \geq 0$. Since $u_0 > 0$, we must have $\alpha(x) \geq 0$. It follows that System 1 has no solution.

Proof of Strong Duality Theorem.

Let $\mu = \inf\{f(x) : x \in X,\ g(x) \leq 0,\ h(x) = 0\}$. If $\mu = -\infty$, then by a corollary of weak duality theorem, $\sup \theta(u, v) :\ u \geq 0\} = -\infty$. So, let us consider the case, μ is finite. Consider the system:

$$f(x) - \mu < 0,\ g(x) \leq 0,\ h(x) = 0,\ x \in X.$$

By the definition of μ, this system has no solution. Hence, from the Lemma, there exists a nonzero vector $(u_0, u, v) \in \mathbf{R}^{1+m+l}$ with $(u_0, u) \geq 0$ such that

$$u_0(f(x) - \mu) + u^t g(x) + v^t h(x) \geq 0\ \forall\ x \in X \tag{42}$$

We first show that $u_0 > 0$. To the contrary, assume $u_0 = 0$. From the hypothesis of the theorem, that $z \in X$ satisfies $g(z) < 0$ and $h(z) = 0$. Substituting z in the above inequality, we get $u^t g(z) \geq 0$. This implies, as $u \geq 0$ and $g(z) < 0$, $u = 0$. From (17), it follows $v^t h(x) \geq 0\ \forall x \in X$. Since $0 \in int(h(X))$, there exists an $x \in X$ such that $h(x) = -\lambda v$ where λ is a small positive real. This implies, $v^t(-\lambda v) \geq 0$ which in turn implies $v = 0$. Thus, $(u_0, u, v) = 0$ which is a contradiction. Hence, $u_0 > 0$. Without loss of generality, we may assume that $u_0 = 1$ and write

$$f(x) + u^t g(x) + v^t h(x) \geq \mu\ \forall x \in X. \tag{43}$$

This implies

$$\theta(u, v) = \inf\{f(x) + u^t g(x) + v^t h(x) :\ x \in X\} \geq \mu.$$

From weak duality theorem, it follows that

$$\inf\{f(x) : x \in X,\ g(x) \leq 0,\ h(x) = 0\} = \sup\{\theta(u, v) : u \geq 0\}$$

Finally, suppose $\bar{x}$ is an optimal solution to the primal problem, that is, $\bar{x} \in X,\ g(\bar{x}) \leq 0,\ h(x) = 0$ and $f(\bar{x}) = \mu$. Sustituting in (18), we get $\bar{u}^t g(\bar{x}) \geq 0$. Since $\bar{u} \geq 0$ and $g(\bar{x}) \leq 0$, $\bar{u}^t g(\bar{x}) = 0$.

Remark. The constraint qualification that $0 \in int(h(X))$ used in Strong Duality Theorem automatically holds good if $X = R^n$. To see this, note that we may assume without loss of generality that the matrix A defining $h(x)$ is of full row rank. If $y \in \mathbf{R}^m$, then $y = h(x)$ where $x = A^t(AA^t)^{-1}(y + b)$ and hence $h(X) = \mathbf{R}^m$. Therefore, $0 \in int(h(X))$.

Saddle Point Optimality criteria

An important consequence of strong duality theorem is the saddle point optimality criteria. The existence of a saddle point asserts optimal solutions to both the primal and dual problems and that the optimal objective values of the two problems are equal. This does not require any convexity assumptions made in the strong duality theorem. However, under the convexity assumptions one can assert the existence of a saddle point.

Saddle Point Theorem. Let X be a nonempty set in R^n, and let $f : R^n \to \mathbf{R}$, $g : R^n \to \mathbf{R}^m$ and $h : R^n \to \mathbf{R}^l$. Suppose there exist $\bar{x} \in X$ and $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$ so that

$$\phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v}) \leq \phi(x, \bar{u}, \bar{v}) \tag{44}$$

for all $x \in X$, for all $u \geq 0$ and for all v, where $\phi(x, u, v) = f(x) + u^t g(x) + v^t h(x)$. Then, $\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal solutions to the primal and dual problems respectively.

Conversely, suppose that X, f, g are convex and that h is affine (i.e., $h(x) = Ax - b$). Further, assume that there exists a $z \in X$ such that $g(z) < 0$ and $h(z) = 0$, and that $0 \in int(h(X))$. If $\bar{x}$ is optimal solution to the primal problem, then there exists $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$, so that (19) hold true.

Proof. Suppose there exist $\bar{x} \in X$ and $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$ such that (19) hold good. Since

$$f(\bar{x}) + u^t g(\bar{x}) + v^t h(\bar{x}) = \phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v})$$

for all $u \geq 0$ and all $v \in \mathbf{R}^l$, it follows that $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$. Therefore, $\bar{x}$ is a solution to the primal problem. Putting $u = 0$ in the above inequality, it follows that $\bar{u}^t g(\bar{x}) \geq 0$. Since $u \geq 0$ and $g(\bar{x}) \leq 0$, $\bar{u}^t g(\bar{x}) = 0$. From (19), for each $x \in X$, we have

$$\begin{aligned} f(\bar{x}) &= f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) \\ &= \phi(\bar{x}, \bar{u}, \bar{v}) \\ &\leq \phi(x, \bar{u}, \bar{v}) = f(x) + \bar{u}^t g(x) + \bar{v}^t h(x) \end{aligned} \tag{45}$$

Since (20) holds good for all $x \in X$, it follow that $f(\bar{x}) \leq \theta(\bar{u}, \bar{v})$. Since $\bar{x}$ is feasible to the primal and $\bar{u} \geq 0$, from a corollary to the weak duality theorem it follows that $\bar{x}$ and $\bar{u}^t g(\bar{x}) = 0$. are optimal to the primal and the dual problems respectively.

Conversely, suppose that $\bar{x}$ is an optimal solution to the primal problem. By strong duality theorem, there exists $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$ such that $f(\bar{x}) = \theta(\bar{u}, \bar{v})$ and $\bar{u}^t g(\bar{x}) = 0$.

By definition of θ, we must have

$$f(\bar{x}) = \theta(\bar{u}, \bar{v}) \leq f(x) + \bar{u}^t g(x) + \bar{v}^t h(x) \forall x \in X$$

But since $\bar{u}^t g(\bar{x}) = 0$,

$$\phi(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) \leq \phi(x, \bar{u}, \bar{v}) \forall x \in X.$$

Again,

$$\begin{aligned}\phi(\bar{x}, u, v) &= f(\bar{x}) + u^t g(\bar{x}) + v^t h(\bar{x}) \\ &\leq f(\bar{x}) = \phi(\bar{x}, \bar{u}, \bar{v})\end{aligned}$$

Thus, $\phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v}) \leq \phi(x, \bar{u}, \bar{v})$.

Relationship Between Saddle Point Criteria and Kuhn-Tucker Conditions

Theorem. Let $S = \{x \in X : g(x) \leq 0,\ h(x) = 0\}$, and consider the primal problem, minimize $f(x)$ subject to $x \in S$. Suppose that $\bar{x} \in S$ satisfies the Kuhn-Tucker conditions, that is, there exist $\bar{u} \geq 0$ and $\bar{v}$ such that

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{u} + \nabla h(\bar{x})\bar{v} = 0,\ \bar{u}^t g(\bar{x}) = 0 \tag{46}$$

Suppose that $f,\ g_i, i \in I$ are convex at $\bar{x}$, where $I = \{i : g_i(\bar{x}) = 0\}$. Further suppose that if $\bar{v}_i \neq 0$, then h_i is affine. Then, $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point, that is,

$$\phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v}) \leq \phi(x, \bar{u}, \bar{v})$$

for all $x \in X$, for all $u \geq 0$ and for all v, where $\phi(x, u, v) = f(x) + u^t g(x) + v^t h(x)$.

Conversely, suppose that $(\bar{x}, \bar{u}, \bar{v})$, with $\bar{x} \in int(X)$ and $\bar{u} \geq 0$, is a saddle point. Then $\bar{x}$ is feasible to the primal problem and furthermore, $(\bar{x}, \bar{u}, \bar{v})$ satisfies the Kuhn-Tucker conditions given in (21).

Proof. Suppose that $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in S$ and $\bar{u} \geq 0$ satisfy Kuhn-Tucker conditions, (21). By convexity of $f,\ g_i, i \in I$ at $\bar{x}$, and since h_is are affine for $v_i \neq 0$, we have

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) \tag{47}$$

$$g_i(x) \geq g_i(\bar{x}) + \nabla g_i(\bar{x})^t (x - \bar{x}) \text{ for } i \in I \tag{48}$$

$$h_i(x) = h_i(\bar{x}) + \nabla h_i(\bar{x})^t (x - \bar{x}) \text{ for } v_i \neq 0 \tag{49}$$

for all $x \in X$. Multiplying (23) by $\bar{u}_i \geq 0$, (24) by $\bar{v}_i$ and adding, and using the hypothesis (20), it follows that $\phi(\bar{x}, \bar{u}, \bar{v}) \leq \phi(x, \bar{u}, \bar{v})$ for all $x \in X$.

Since $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{u}^t g(\bar{x}) = 0$, it follows that $\phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v})$ for all $u \geq 0$. Hence, $(\bar{x}, \bar{u}, \bar{v})$ satisfies the saddle point condition.

To prove the converse, suppose that $(\bar{x}, \bar{u}, \bar{v})$, with $\bar{x} \in int(X)$ and $\bar{u} \geq 0$, is a saddle point. Since $\phi(\bar{x}, u, v) \leq \phi(\bar{x}, \bar{u}, \bar{v})$ for all $u \geq 0$ and all v, it follows $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{u}^t g(\bar{x}) = 0$. This shows that $\bar{x}$ is a feasible solution to the primal. Since $\phi(\bar{x}, \bar{u}, \bar{v}) \leq \phi(x, \bar{u}, \bar{v})$ for all $x \in X$, $\bar{x}$ is a local optimal solution to the problem: minimize $\phi(x, \bar{u}, \bar{v})$ subject to $x \in X$. Since $\bar{x} \in int(X)$, $\nabla_x \phi(\bar{x}, \bar{u}, \bar{v}) = 0$, that is, $\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{u} + \nabla h(\bar{x})\bar{v} = 0$. It follows that (21) holds good.

Remark. We see that under certain convexity assumptions, the Lagrangian multipliers of Kuhn-Tucker conditions also serve as the multipliers in the saddle point criteria. Conversely, the multipliers of the saddle point criteria are the Lagrangian multipliers of the Kuhn-Tucker conditions. Also, note that the dual varibles turn out to be the Lagrangian multipliers.

Properties of the Dual Function

For the problems with zero duality gap, one way of solving the primal problem is to obtain the solution via the dual problem. In order to solve the dual problem one has to understand the properties of the dual objective function. We shall derive some properties of the dual under assumption that the set X is a compact set. As one can always impose boundary conditions on the variables x, this assumption is a reasonable one to make.

For ease of notation, we shall combine the vector functions g and h into β, i.e., $\beta(x) = (g(x)^t, h(x)^t)^t$ and combine the dual variable vectors u and v into w, i.e., $w = (u^t, v^t)^t$.

The first property of the dual objective function is that it is concave over the entire $\mathbf{R}^{m+l}$ which in turn asserts that any local optimal solution is global optimal solution to the dual maximization objective.

Theorem. Let X be a nonempty compact set in R^n. Let $f : R^n \to \mathbf{R}$, and $\beta : R^n \to \mathbf{R}^{m+l}$ be continuous. Then, θ defined by

$$\theta(w) = \inf\{f(x) + w^t\beta(x) : \ x \in X$$

is concave over $\mathbf{R}^{m+l}$.

Proof. Since X is compact and since f and β are continuous, θ is a real valued function on $\mathbf{R}^{m+l}$. For any $w_1, \ w_2 \in \mathbf{R}^{m+l}$ and for any $\lambda \in (0, 1)$, we have

$$\begin{aligned}
&\theta[\lambda w_1 + (1-\lambda)w_2] \\
&= \inf\{f(x) + [\lambda w_1 + (1-\lambda)w_2]^t\beta(x) : \ x \in X\} \\
&= \inf\{\lambda[f(x) + w_1^t\beta(x)] + (1-\lambda)[f(x) + w_2^t\beta(x)] : x \in X\} \\
&\geq \lambda \inf\{f(x) + w_1^t\beta(x) : x \in X\} \\
&\qquad + (1-\lambda)\inf\{f(x) + w_2^t\beta(x) : x \in X\} \\
&= \lambda\theta(w_1) + (1-\lambda)\theta(w_2).
\end{aligned}$$

When X is compact and the f and β are continuous, the infimum defined by $\theta(w) = \inf\{f(x) + w^t\beta(x) : x \in X\}$ is attained at some $x \in X$ for each w. We shall define the set $X(w) = \{x \in X : f(x) + w^t\beta(x) = \theta(w)\}$. If $X(w)$ is a singleton set, then θ is differentiable at w.

Lemma. Let X be a nonempty compact set in R^n and let $f : R^n \to \mathbf{R}$, $\beta : R^n \to \mathbf{R}^{m+l}$ be continuous. Let $\bar{w} \in \mathbf{R}^{m+l}$ be such that $X(\bar{w})$ is a singleton, say $\{\bar{x}\}$. If w^k is any sequence such that $w^k \to \bar{w}$, then for any sequence x^k, with $x^k \in X(w^k)$ for each k, converges to $\bar{x}$.

Proof. Suppose x^k does not converge to $\bar{x}$. Since X is compact, we may assume without loss of generality that x^k converges to $z \in X$ where $z \neq \bar{x}$. For each k, as $x^k \in X(w^k)$

$$f(x^k) + \beta(x^k)^t w^k \leq f(\bar{x}) + \beta(\bar{x})^t w^k$$

Taking the limit as $k \to \infty$, we get $f(z) + \beta(z)^t\bar{w} \leq f(\bar{x}) + \beta(\bar{x})^t\bar{w}$. This implies $z \in X(\bar{w}) = \{\bar{x}\}$, contradiction.

Theorem. Let X be a nonempty compact set in R^n and let $f : R^n \to \mathbf{R}$, $\beta : R^n \to \mathbf{R}^{m+l}$ be continuous. Let $\bar{w} \in \mathbf{R}^{m+l}$ be such that $X(\bar{w})$ is a singleton. Then θ is differentiable at $\bar{w}$ with gradient $\nabla\theta(\bar{w}) = \beta(\bar{x})$.

Proof. Since f and β are continuous, and X is compact, for any w there exists $x_w \in X(w)$. From the definition of θ, the following inequalities hold good:

$$\begin{aligned} \theta(w) - \theta(\bar{w}) &\leq f(\bar{x}) + w^t\beta(\bar{x}) - f(\bar{x}) + \bar{w}^t\beta(\bar{x}) \\ &= (w - \bar{w})^t\beta(\bar{x}) \end{aligned} \tag{50}$$

$$\begin{aligned} \theta(\bar{w}) - \theta(\bar{w}) &\leq f(x_w) + \bar{w}^t\beta(x_w) - f(x_w) - w^t\beta(x_w) \\ &= (\bar{w} - w)^t\beta(x_w) \end{aligned} \tag{51}$$

From (25) and (26) and Schwartz inequality, it follows that

$$\begin{aligned} 0 &\geq \theta(w) - \theta(\bar{w}) - (w - \bar{w})^t\beta(\bar{x}) \\ &\geq (w - \bar{w})^t[\beta(x_w) - \beta(\bar{x})] \\ &\geq - \mid\mid w - \bar{w} \mid\mid \, \mid\mid \beta(x_w) - \beta(\bar{x}) \mid\mid \end{aligned}$$

This further implies that

$$0 \geq \frac{\theta(w) - \theta(\bar{w}) - (w - \bar{w})^t\beta(\bar{x})}{\mid\mid w - \bar{w} \mid\mid} \geq - \mid\mid \beta(x_w) - \beta(\bar{x}) \mid\mid \tag{52}$$

As $w \to \bar{w}$, then by by Lemma and by continuity of β, $\beta(x_w) \to \beta(\bar{x})$. From (27), we get

$$\lim_{w \to \bar{w}} \frac{\theta(w) - \theta(\bar{w}) - (w - \bar{w})^t\beta(\bar{x})}{\mid\mid w - \bar{w} \mid\mid} = 0$$

Hence θ is differentiable at $\bar{w}$ with gradient $\beta(\bar{x})$.