IME 637

Some example LPPs

Q.

	Chairs	Tables	Desks	Available
Profit	50	80	90	
Labour	9	7	8	200
Raw Material	8	11	9	400

Let no. of C, T, D be x_1, x_2, x_3

Objective func: max $Z = 50x_1 + 80x_2 + 90x_3$

Constraints: $\begin{cases} 9x_1 + 7x_2 + 8x_3 \leq 200 \\ 8x_1 + 11x_2 \quad 9x_3 \leq 400 \end{cases}$

Non-negativity constraint ← $x_1, x_2, x_3 \geq 0$

x_1 → Decision Variables

General Form: Max / Min $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

s.t.: $a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq | \geq | = b_1$

$\vdots$

$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq | \geq | = b_m$

& $x_1, x_2, \ldots x_n \geq 0$

Q1 s, t max. $z = 3s + 2t$ s.t. $2s + t \leq 100$, $s + t \leq 80$, $s \leq 40$, $s \geq 0$, $t \geq 0$

Q2 t_1, t_2 max. $z = 300t_1 + 500t_2$ s.t. $\frac{t_1}{800} + \frac{t_2}{700} \leq 1$, $t_1 \geq 0$, $t_2 \geq 0$

$t_1/1500 + t_2/1200 \leq 1$

Q3 s_1, s_2, s_3 : min. $z = 5s_1 + 4s_2 + 3s_3$ s.t.

$0 \leq s_1, s_2, s_3 \leq 700$, $s_1 \cdot 40\% + s_2 \cdot 30\% + s_3 \cdot 20\% \geq 500$

$s_1 \cdot 40\% + s_2 \cdot 35\% + s_3 \cdot 20\% \geq 300$

$s_1 \cdot 20\% + s_2 \cdot 35\% + s_3 \cdot 60\% \geq 300$

Q4 $0 \leq m, t^u, w, t^n, f, s^a, s^u$: min $z = m + t^u + w + t^n + f + s^a + s^u$

$m + t^n + f + s^a + s^u \geq 17$, $f + s^a + s^u + m + t^u \geq 13$, ...

Course Outline

- IPP: Integer Programming Problem
- DP: Dynamic Programming
- NLP: Non-Linear Programming
- MultiCriteria Decision Making
- AHP: Analytical Heirarchy Process
- Queuing Models
- DEA: Data Envelopment Analysis

Integer Programming Problem

An IPP is an LPP in which some or all of the variables are required to be non-negative integers.

- An IPP in which all variables required to be integers is called a pure IPP.
- If only some variables required to be integers then it is called a mixed IPP.
- If all variables are required to be either 0 or 1 then it is called a 0-1 IPP or a binary IPP.

Def The LPP obtained by omitting all integer or 0-1 constraints on variables is called LP relaxation of the LPP.

★ Solving the LP relaxation and rounding-off does not always directly give the solution.
However, it does provide insights to the original LPP.

Assignment Problem: There are n people available to carry out n jobs. Each person is assigned to carry out exactly one job. Every job has to be completed. There is an estimated cost of c_{ij} if person i is assigned to job j.

The objective is to find minimum cost of assignment.

$$x_{ij} = \begin{cases} 1, & \text{if } i^{th} \text{ person is assigned to } j^{th} \text{ job} \\ 0, & \text{otherwise} \end{cases}$$

Person \ Job	1	2	3	...	n
1	c_{11}	c_{12}	c_{13}	...	c_{1n}
2	c_{21}	c_{22}	...	...	c_{2n}
3	⋮	⋮	⋮		⋮
⋮	⋮		⋮		⋮
n	c_{n1}	c_{n2}	...	...	c_{nn}

Formulation –
- Objective function: $\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \cdot x_{ij}$
- Constraints:

Each person assigned exactly one job.
$$\begin{cases} x_{11} + x_{12} + \ldots + x_{1n} = 1 \\ x_{21} + x_{22} + \ldots + x_{2n} = 1 \\ \vdots \\ x_{n1} + x_{n2} + \ldots + x_{nn} = 1 \end{cases}$$

Each job assigned to exactly one person
$$\begin{cases} x_{11} + x_{21} + \ldots + x_{n1} = 1 \\ x_{12} + x_{22} + \ldots + x_{n2} = 1 \\ \vdots \\ x_{1n} + x_{2n} + \ldots + x_{nn} = 1 \end{cases}$$

$x_{ij} = 0$ or 1 $\forall\, i,j$

0-1 / Binary Constraint

Travelling Salesman Problem (TSP)

A salesman must visit each n cities exactly once and then return to his starting point. The time taken to travel from city i to city j is c_{ij}

The objective is to find the order in which he should make his tour.

$$x_{ij} = \begin{cases} 1, & \text{city } j \text{ visited from city } i \\ 0, & \text{otherwise} \end{cases}$$

Obj. function: $\min. \sum_{i,j=1}^{n} c_{ij} \cdot x_{ij}$

Constraints: $\underbrace{\sum_{i=1}^{n} x_{ij} = 1}_{j^{th} \text{ city entered exactly once}} = \underbrace{\sum_{i=1}^{n} x_{ji}}_{j^{th} \text{ city exited exactly once}} \quad \forall \ 1 \le j \le n$

$\underbrace{x_{ij} \in \{1, 0\}}_{0-1 \text{ Constraint}} \ \forall \ 1 \le i, j \le n; \quad \underbrace{\sum_{i \in S} \sum_{j \notin S} x_{ij} \ge 1}_{\text{single path connects all } n \text{ points}} \ \forall \ \phi \ne S \subset [n]$

Either - Or Constraint

Consider the following constraint $f(x_1, x_2 \dots x_n) \le 0$, $g(x_1, x_2, \dots, x_n) \le 0$. Suppose we want at least one of the above constraints to be satisfied.

These are called either-or constraints.

☆ Let y be a 0-1 variable, i.e; y takes value 0 or 1. Let M be a very large positive value.

$y \in \{0, 1\}, \quad f(x_1 \dots x_n) \le My, \quad g(x_1 \dots x_n) \le M(1-y)$

Q. A car company is considering manufacturing 3 types of cars: L, M, S. The resources required for & the profits yielded for each type of car are:

Currently, 6000 tons of steel & 60,000 hrs. of labour are available.

	S	M	L
Steel	1.5 tons	3 tons	5 tons
Labour	30 hrs.	25	40
Profit	2000	3000	4000

For a production of a car type to be economically feasible, at least 1000 cars of that type must be produced. Formulate an IPP to maximise the company's profit.

A: $s, m, l \equiv x_1, x_2, x_3$

$\max (2x_1 + 3x_2 + 4x_3) \cdot 1000$ s.t. $1.5x_1 + 3x_2 + 5x_3 \le 6000$,

$30x_1 + 25x_2 + 40x_3 \le 60{,}000$, $0 \le x_1, x_2, x_3 \in \mathbb{N}$

$(x_1 \ge 1000 \text{ OR } x_1 \le 0) \equiv 1000 - x_1 \le M_1 y_1, \quad x_1 \le M_1(1-y_1)$

||ly, $1000 - x_2 \le M_2 y_2$, $x_2 \le M_2(1-y_2)$, $1000 - x_3 \le M_3 y_3$, $x_3 \le M_3(1-y_3)$
s.t. M_i = large +ve const., $y_1, y_2, y_3 \in \{0,1\}$

NOT decision variables

Can use the same M in place of all the M_i but using different M_i's allows for stricter bounds

For eg. choose $M_1 = M_2 = 2000$, $M_3 = 1200$

If-Then Constraint

Suppose that we want to ensure that if a constraint $f(x_1 \dots x_n) > 0$ is satisfied then the constraint $g(x_1 \dots x_n) \geq 0$ must also be satisfied. But, if $f(x_1 \dots x_n) > 0$ is not satisfied then $g(x_1 \dots x_n) \geq 0$ may or may not be satisfied.

☆ Let $p := (f(x_1 \dots x_n) > 0)$ & $q := (g(x_1 \dots x_n) \geq 0)$

Now, $p \Rightarrow q$ is the same as $\neg p \vee q$

So, we just want either $(f \le 0)$ OR $(g \geq 0)$ $(-g \le 0)$

$f \le My$, $g \geq M(y-1)$

Set Covering Problem

Given a certain number of regions, the problem is to decide where to install a set of emergency service centers. The cost of installing a service center & which regions it can service are known. The objective is to choose a set of service centers so that each region is covered at a minimum cost.

Let $M = \{1, 2, \dots, m\}$ be the set of regions, $N = \{1, 2, \dots, n\}$ be the set of service centers, c_j be the cost of installing service center $j \in N$, and matrix $A_{m \times n}$ be the incidence matrix with entries,

$$A_{ij} = \begin{cases} 1, & \text{region } i \text{ is covered by center } j \\ 0, & \text{otherwise} \end{cases}$$

$x_j = 1$ if center j is installed, & 0 otherwise

$$\min \sum_{j \in N} c_j \cdot x_j \quad \text{s.t.} \quad \sum_{j \in N} A_{ij} \cdot x_j \geq 1 \ \forall i \in M \ \& \ x_j \in \{0,1\} \ \forall j \in N$$

0-1 Knapsack Problem

There is a budget b available for investment in project during the coming year & n projects are under consideration, a_j is the outlay for j^{th} project & c_j is the expected return.

Formulate an IPP which chooses a set of projects so that the budget is not exceeded & expected return is maximised.

$x_j = 1$ if j^{th} project is undertaken, 0 otherwise

$$\max \sum_{j \in [n]} x_j \cdot c_j \quad \text{s.t.} \quad \sum_{j \in [n]} x_j \cdot a_j \leq b$$

Solving an LPP

Q $\quad x_1 + x_2 = 3$

$-x_2 + x_3 = -1$

$n = 3,\ m = 2,\ n - m = 1$

$\longrightarrow$

$x_1 = 0$ } NBV $(= n - m)$

$x_2 = 3$, $x_3 = 2$ } BV $(= m)$

NBV: Non-Basic Variables; Equal to zero

BV: Basic Variable; Not equal to zero

BFS: Basic Feasible Solution; Solution with minimal set of non-zero variables.

☆ The set of BV's forms the 'Basis'

Q. Max $z = 60x_1 + 30x_2 + 20x_3$, $\quad x_1, x_2, x_3 \geq 0$

$8x_1 + 6x_2 + x_3 \leq 48$

$4x_1 + 2x_2 + 1.5x_3 \leq 20$

$$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$$
$$x_2 \leq 5$$

Add slack variables & make them basic

$$\begin{aligned}
Z - 60x_1 - 30x_2 - 20x_3 &= 0 \\
8x_1 + 6x_2 + x_3 + s_1 &= 48 \\
4x_1 + 2x_2 + 1.5x_3 + s_2 &= 20 \\
2x_1 + 1.5x_2 + 0.5x_3 + s_3 &= 8 \\
x_2 + s_4 &= 5
\end{aligned}$$

$x_1 = x_2 = x_3 = 0$, $s_1 = 48$, $s_2 = 20$, $s_3 = 8$, $s_4 = 5$ | BFS

- Choosing x_1 due to most -ve co-eff. in Z-equation, $s_1 = 48 - 8x_1 \geq 0 \Rightarrow x_1 \leq 6$, s_2: $x_1 \leq 5$, s_3: $x_1 \leq 4$, s_4: -
- $\therefore$ x_1 is the entering variable, s_3 is the exiting variable (from basis), $x_1 = 4$, x_2, x_3 remain zero
- Make co-efficient of x_1 in R_4 equal to 1: $R_4 \leftarrow R_4/2$
- Eliminate x_1 from all the other rows: $R_1 \leftarrow R_1 + 60R_4$, $R_2 \leftarrow R_2 - 8R_4$, $R_3 \leftarrow R_3 - 4R_4$

EROs
Elementary Row Operations

- Repeat these until z can no longer increase. In a max. problem, stop when Z equation has no -ve co-effs. left.
- This is the **Simplex Method** to solve LPPs.

☆ If the inequality is $\sum a_i x_i \leq c_0$, add a **slack variable** s_0 on the LHS, $\sum a_i x_i + s_0 = c_0$
If the inequality is $\sum a_i x_i \geq c_0$, add an **excess/surplus variable** e_0 on the RHS, $\sum a_i x_i = c_0 + e_0$. In general, just set

$$\sum a_i x_i \leq c_0 \Rightarrow 0 \leq c_0 - \sum a_i x_i =: s_0 \Rightarrow c_0 = \sum a_i x_i + s_0$$
$$\sum a_i x_i \geq c_0 \Rightarrow e_0 := \sum a_i x_i - c_0 \geq 0 \Rightarrow e_0 + c_0 = \sum a_i x_i$$

☆ If the problem is about minimization, choose the variable with the most positive co-efficient or just maximize the negative of the obj. function

Q. max $z = 60x_1 + 35x_2 + 20x_3$

s.t.
$$8x_1 + 6x_2 + x_3 \le 48$$
$$4x_1 + 2x_2 + 1.5x_3 \le 20$$
$$2x_1 + 1.5x_2 + 0.5x_3 \le 8$$
$$x_2 \le 5, \quad x_1, x_2, x_3 \geq 0$$

Final Table:

	Z	x_1	x_2	x_3	s_1	s_2	s_3	s_4	=
Z	1					10	10		280
s_1			-2		1	2	-8		24
x_3			-2	1		-2	-4		8
x_1		1	1.25			-0.5	1.5		2
s_4			1					1	5
Z	1					10	10		280
s_1		1.6			1	1.2	-5.6		27.2
x_3		1.6		1		1.2	-1.6		11.2
x_2		0.8	1			-0.4	1.2		1.6
s_4		-0.8				0.4	-1.2	1	3.4

x_1	x_2	x_3
2	0	8

x_2 can enter

$\Rightarrow x_1$ exits

$\downarrow$

Multiple Solutions

Q. max $z = 36x_1 + 30x_2 - 3x_3 - 4x_4$

s.t.
$$x_1 + x_2 - x_3 \le 4$$
$$6x_1 + 5x_2 - x_4 \le 10$$

Final Table:

	Z	x_1	x_2	x_3	x_4	s_1	s_2	=
Z	1		2	-9		12	4	100
x_4			1	-6	1	6	-1	20
x_1		1	1	-1		1		5

- Look at x_3 : $x_3 \uparrow \Rightarrow z \uparrow$

  But rows R_2, R_3 don't put a constraint on x_3. So, we can $\uparrow$ x_3 indefinitely & z will keep increasing with it.

Q. A cloth company is capable of manufacturing 3 types of clothing shirts, shorts & pants. The manufacture of each type of clothing requires an appropriate type of machinery. The machinery needed to manufacture each type of clothing must be rented at the following rates: shirt 200/week, shorts 150/wk, pants 100/wk. The manufacture of each type of clothing & labour as shown:

	Labour (hrs.)	Cloth (sq. yards)
Shirts	3	4
Shorts	2	3
Pants	6	4

Each week, 150 hrs. of labour & 160 sq. yards of cloth are available.

The variable unit cost & selling price for each type of clothing is:

	Sales Price	Variable Cost
Shirts	12	6
Shorts	8	4
Pants	15	8

Formulate an IPP whose solution will maximize the company's weekly profits.

A: max. $(12-6)x_1 + (8-4)x_2 + (15-8)x_3 - 200y_1 - 150y_2 - 100y_3$

s.t. $3x_1 + 2x_2 + 6x_3 \leq 150$, $4x_1 + 3x_2 + 4x_3 \leq 160$,

$\mathbb{Z} \ni x_1, x_2, x_3 \geq 0$, $y_1, y_2, y_3 \in \{0,1\}$,

$\rightarrow y_1 := (x_1 > 0)$, ie, If $(x_1 > 0)$ then $y_1 = 1$

$\Rightarrow (x_1 \leq 0)$ OR $(y_1 \geq 1) \Rightarrow x_1 \leq M_1 y_1$

& $x_2 \leq M_2 y_2$, $x_3 \leq M_3 y_3$ // $(x_1 > 0) \Rightarrow (y_1 = 1)$

{ If $x_i = 0$, then $y_i = 0$ or $y_i = 1$, both possible. But, obj. function would force y_i to be 0 }

Solving IPPs

Implicit Enumeration (To solve 0-1 IPP) : Brute-Force with some heuristics and early stopping.

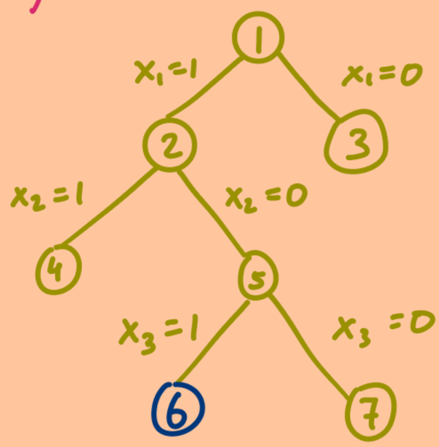
Q max $z = -7x_1 - 3x_2 - 2x_3 - x_4 - 2x_5$

s.t. $-4x_1 - 2x_2 + x_3 - 2x_4 - x_5 \le -3$

$-4x_1 - 2x_2 - 4x_3 + x_4 + 2x_5 \le -7$

$x_i \in \{0, 1\}\Big|_{i=1}^{i=5}$

S:
- Each $x_i = 0$ is the best completion but constraints are not satisfied. However, each constraint can be satisfied individually.
- At node 1, all vars. are free vars.
- At node 2, $x_1 = 1$, best completion not possible. But $\exists$ a feasible solⁿ
- At node 5, $x_1 = 1$, $x_2 = 0$, best completion is $x_3 = x_4 = x_5 = 0$, does not satisfy 2nd constraint. But, $\exists$ a feasible solⁿ.



- At node 6, best completion $x_4 = x_5 = 0$ satisfies the constraints. $z = -9$. Node 6 will be fathomed.
- Going deeper from node 6, will yield sub-optimal solutions. So, stop at node 6. For other nodes, use $z = -9$ as the benchmark.
- If at some other node, best completion not better than benchmark, mark that node infeasible and remove it from consideration

for further branching.

- At node 7, z value is better, but 2^{nd} constraint is not satisfied even by the constraint's best completion. ∴ Stop branching from node 7. Mark it infeasible ∵ no feasible completion.
- At node 4, best completion yields $z = -10$, it cannot beat current benchmark. Stop branching.
- At node 3, best completion is $(0, 1, 1, 1, 1) = -8$, but 2^{nd} constraint cannot be satisfied even by its best completion. ∴ Stop branching.

Cutting Plane Algorithm (To solve pure IPP): A variant can be used to solve mixed IPPs too.

Q. max. $z = 8x_1 + 5x_2$ s.t. $x_1 + x_2 \le 6$,
$9x_1 + 5x_2 \le 45$, $x_1, x_2 \ge 0$ & integers

- Solve the LP relaxation of the IPP.
  If the optimal solution of the LP relaxation has all integer values then it is also optimal for the IPP & we stop.

	z	x_1	x_2	s_1	s_2	RHS
z	1	−8	−5			0
s_1		1	1	1		6
s_2		9	5		1	45
z	1			1.25	0.75	41.25
x_2			1	2.25	−0.25	2.25
x_1		1		−1.25	0.25	3.75

- Optimal sol^n of the LP relaxation is:

$x_1 = 3.75, \quad x_2 = 2.25, \quad z = 41.25$

- Let $[x]$ is the largest integer $\le x$.
  $[3.99] = 3, \quad [-1.25] = -2, \quad x = [x] + f, \quad 0 \le f < 1$
- Choose a Source Row, say R_2 (0-based).
  Now, it can be re-written as:

$$x_1 - 2s_1 + 0.75s_1 + 0.25s_2 = 3 + 0.75$$

- Rewrite again by taking all integer terms to the LHS and fractional terms to the RHS.

$$x_1 - 2s_1 - 3 = 0.75 - 0.75s_1 - 0.25s_2$$

- Define the cut as follows:

$$0.75 - 0.75s_1 - 0.25s_2 \le 0, \text{ i.e., RHS} \le 0$$

Cut can be written as: $-0.75s_1 - 0.25s_2 \le -0.75$

- Add this constraint to the optimal simplex table of LP relaxation & apply dual simplex method. The new table is:

	z	x_1	x_2	s_1	s_2	s_3	RHS
z	1			1.25	0.75		1.25
x_2			1	2.25	-0.25		2.25
x_1		1		-1.25	0.25		3.75
$\leftarrow s_3$				-0.75	-0.25	1	-0.75

RHS is negative

-ve co-efficients: Compare absolute value of ratios w/ z row

min. $\{1.25/0.75, \ 0.75/0.25\}$

$\therefore$ Choose s_1

	z	x_1	x_2	s_1	s_2	s_3	RHS
z	1			1.25	0.75		1.25
x_2			1	2.25	-0.25		2.25
x_1		1		-1.25	0.25		3.75
$\rightarrow s_1$				1	$1/3$	$-1/0.75$	1

- Eliminate s_1 from other rows using the last row.

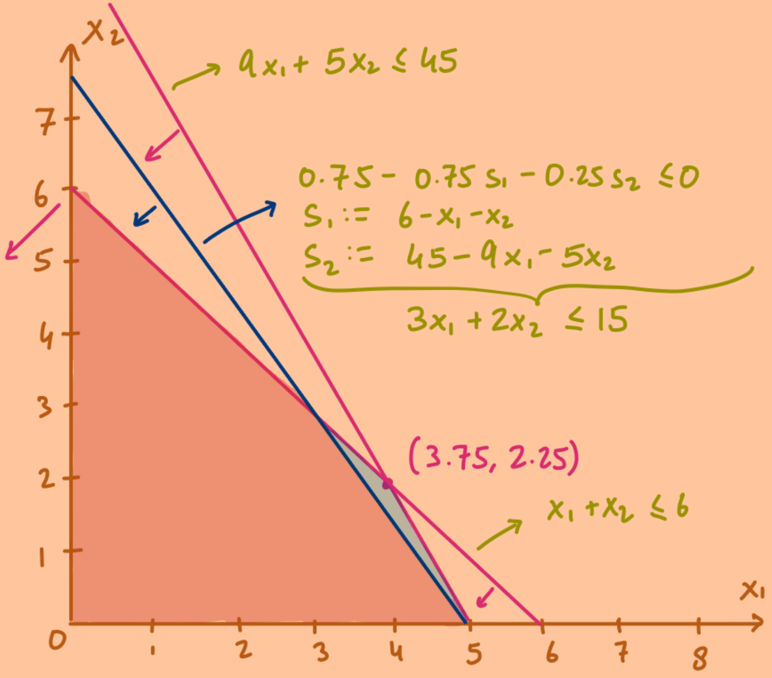
	z	x_1	x_2	s_1	s_2	s_3	RHS
z	1				$1/3$	$5/3$	40
x_2			1		-1	3	0
x_1		1			$2/3$	$-5/3$	5
s_1				1	$1/3$	$-4/3$	1

- $x_2 = 0$, $x_1 = 5$, $s_1 = 1$, $z = 40$; All x_i's are integer, the RHS contains only integers and it is a feasible solution
  $\therefore$ This is an optimal solution.
- The used cut is known as Gomory's Cut.

☆ LHS of the re-written Source Row is integer $\Rightarrow$ RHS is integer.
Cut from this Source Row has 2 properties:
1) Cut preserves all points of the IPP.
2) Cut eliminates the optimal solution of the LP relaxation.

- Now, the RHS of the Source Row:
  $0.75 - 0.75\, s_1 - 0.75\, s_2$ is an integer. Also, $s_1, s_2 \geq 0 \Rightarrow$ The value is $\leq 0.75 < 1$
  $\therefore$ writing the inequality ≤ 0 preserves the feasible region of the original IPP, $\because$ ≤ 0.75 & being an integer leaves ≤ 0 as the only option.
- $\because$ Source Row would have had only 1 B.V., all the variables on the RHS would be N.B.V.s, i.e., 0. Thus, the optimal solution to the LP relaxation would never be able to satisfy the cut's inequality. Hence, it is always eliminated in the new table.
- Repeat the procedure until all take integral values.

- The graph of the above problem would look something like:



Cutting Plane Method for mixed IPPs

- Solve the LP relaxation of the mixed IPP. If variables which are required to be integers also happen to be integers in the optimal table, then we have obtained optimal solution of the mixed IPP.
- Otherwise, choose a Source Row (w/ B.V. $\in \mathbb{Z}$, but R.H.S. $\notin \mathbb{Z}$):

$$x_k + \sum_{j \in NB} \alpha_{k_j} w_j = \beta_k = [\beta_k] + f_k$$
$$\Rightarrow x_k - [\beta_k] = f_k - \sum_{j \in NB} \alpha_{k_j} w_j$$

- Now, we want either $x_k \leq [\beta_k]$ OR $x_k \geq [\beta_k] + 1$

$$\therefore \quad f_k - \sum_{j \in NB} \alpha_{k_j} w_j \leq 0 \quad \text{OR} \quad f_k - \sum_{j \in NB} \alpha_{k_j} w_j \geq 1$$

$$\Rightarrow \text{Either} \quad -\sum_{j \in NB} \alpha_{k_j} w_j \leq -f_k \quad \text{OR} \quad \sum_{j \in NB} \alpha_{k_j} w_j \leq f_k - 1$$

Define: $J^+ := \{ j \mid \alpha_{k_j} > 0 \}$, $J^- := \{ j \mid \alpha_{k_j} < 0 \}$

$$-\sum_{j \in J^+} \alpha_{k_j} w_j \leq -\sum_{j \in NB} \alpha_{k_j} w_j \leq -f_k \quad \text{(OR)} \quad \sum_{j \in J^-} \alpha_{k_j} w_j \leq \sum_{j \in NB} \alpha_{k_j} w_j \leq f_k - 1$$

$$\sum_{j \in J^-} \alpha_{k_j} w_j \, f_k / (1 - f_k) \leq -f_k$$

Multiply with $f_k / (1 - f_k) \geq 0$

$$\left.\begin{array}{l} \text{Now,}\quad \alpha_{k_j}\Big|_{j\in J^+} > 0 \Rightarrow -\sum_{j\in J^+} \alpha_{k_j} w_j \le 0 \\ \alpha_{k_j}\Big|_{j\in J^-} < 0,\ f_k/1-f_k \geq 0 \Rightarrow \sum_{j\in J^-} \alpha_{k_j} w_j \frac{f_k}{1-f_k} \le 0 \end{array}\right\} \text{Always}$$

∴ We would like to add the following constraint:

$$\sum_{j\in J^-} \alpha_{k_j} w_j \frac{f_k}{1-f_k} - \sum_{j\in J^+} \alpha_{k_j} w_j \le -f_k$$

This should always hold.

Add a slack variable, add the constraint to optimal table and apply Dual Simplex.

Called the **m-cut**.

- This m-cut can be improved to an m^*-cut so that there are lesser number of iterations.
- For the same, note that some of the w_j's may be integer variables. Let α_{k_j} be the corresponding co-efficient of such w_j's:

$$\alpha_{k_j} = [\alpha_{k_j}] + f_{k_j},\quad 0 \le f_{k_j} < 1$$

Use fractional part with the same sign as α_{k_j} ∴ For those j s.t. $w_j \in \mathbb{Z}$, replace α_{k_j} by f_{k_j} if $\alpha_{k_j} > 0$ & α_{k_j} by $(f_{k_j} - 1)$ if $\alpha_{k_j} < 0$ in the m-cut to get the new cut.

☆ The source row eq^n shows that the integer property of x_k will not be disturbed if α_{k_j} associated with an integer is increased or decreased by an integer.

For an integer w_j, the smallest value of $\alpha_{k_j} < 0$ is $f_{k_j} - 1$ & for $\alpha_{k_j} > 0$ is f_{k_j}. In the m-cut, the smallest absolute value for the co-efficients must be $f_k/(1-f_k) \cdot (1-f_{k_j})$ for $j \in J^-$ & f_{k_j} for $j \in J^+$

☆ $\because$ w_j is an integer, adding 1 to the smallest -ve value of α_{k_j} or subtracting 1 from the smallest +ve value of α_{k_j} should not matter.

Let $\lambda_{k_j} = \min\{ f_{k_j},\ f_k(1-f_{k_j})/(1-f_k)\}\ \forall\ j \in I$

$\mathbb{Z}^c := \mathbb{R} \setminus \mathbb{Z}$, $I = \{ j : w_j \in \mathbb{Z}\}$

$J_n^- = \{ j : j \in J^-,\ w_j \in \mathbb{Z}^c\}$,

$J_n^+ = \{ j : j \in J^+,\ w_j \in \mathbb{Z}^c\}$

$$\boxed{\sum_{j \in J_n^-} \frac{f_k}{1-f_k}\alpha_{k_j} w_j - \sum_{j \in J_n^+} \alpha_{k_j} w_j - \sum_{j \in I} \lambda_{k_j} w_j \leq -f_k}$$

Stronger m-cut OR m^*-cut

Ex. $\max z = -5x_2 - 10x_4 + 20$

s.t. $x_1 - 5/3\, x_2 - 1/3\, x_4 = 5/3$

$-4/3\, x_2 + x_3 + 11/3\, x_4 = 7/3$,

$x_1, x_2, x_3, x_4 \geq 0$ & $x_3, x_4 \in \mathbb{Z}$

Optimal table for LP relaxation :

z	x_1	x_2	x_3	x_4	RHS
1		5		10	20
	1	$-5/3$		$-1/3$	$5/3$
		$-4/3$	1	$11/3$	$7/3$

$\therefore z = 20$

$x_1 = 5/3$, $x_3 = 7/3$

Source Row

$f_k = f_2 = \{7/3\} = 1/3$, $\alpha_{2_4} = 11/3 = 3 + 2/3$ due to $x_4 \in \mathbb{Z}$

$\lambda_{2_4} = \min.\{ 1/3 \cdot (1-2/3)/(1-1/3),\ 2/3\} = \min\{1/6, 2/3\} = 1/6$

$\therefore$ m^*-cut : $\dfrac{-4/3 \cdot 1/3}{1-1/3} x_2 - 1/6\, x_4 = -\dfrac{2x_2}{3} - \dfrac{x_4}{6} \leq \dfrac{-1}{3}$

$+\alpha_{22} \cdot \frac{f_k}{1-f_k} x_2 - \lambda_{2_4} x_4 \leq -f_k$

$\Rightarrow -2x_2/3 - x_4/6 + x_5 = -1/3$

New m^*-cut : (x_5) 0, 0, $-2/3$, 0, $-1/6$, $\overset{x_5}{1}$ = $-1/3$

Apply Dual Simplex to choose $\min.\left\{ \overset{\checkmark (x_2)}{\dfrac{5}{2/3}},\ \overset{(x_4)}{\dfrac{10}{1/6}} \right\} \rightarrow x_2$

Apply EROs to obtain this optimal table

z	x_1	x_2	x_3	x_4	x_5	RHS
1	0	0	0	35/4	15/2	35/2
0	1	0	0	1/12	-5/2	5/2
0	0	0	1	4	-2	3
0	0	1	0	1/4	-3/2	1/2

OR: $z = 35/2,\ x_1 = 5/2,\ x_2 = 1/2,\ x_3 = 3,\ x_4 = 0$

Multiattribute Decision Making

There are multiple goals, all of which must be achieved / fulfilled, but there is no function to be optimized "as such".

- Sometimes, there is no feasible region at all.

Goal Programming: Method to solve Multiattribute Decision Making problems.

Ex: x_1, x_2: No. of minutes of Ads during football games / soap operas.

$7x_1 + 3x_2 \geq 40$, $10x_1 + 5x_2 \geq 60$, $5x_1 + 4x_2 \geq 35$,
$100x_1 + 60x_2 \leq 600$ ⟶ just constraints / goals, no objective function

can add a dummy ⟵ : min. $z = 0 \cdot x_1 + 0 \cdot x_2$

- Any point in the feasible region will do the job. But, there are no feasible points.

- Introduce **Deviational Variable** which capture how much we are deviating from the required goal.

s_i^+ : Amount by which the goal is "oversatisfied"
s_i^- : " " " " " " "undersatisfied"

$s_i^+, s_i^- \geq 0$ but must ensure at most $\underbrace{\text{one variable is active}}$ at a time

$7x_1 + 3x_2 \geq 40$ $\quad$ $7x_1 + 3x_2 + s_1^- - s_1^+ = 40$

- Similarly, for other two goals. ∵ $100x_1 + 60x_2 \leq 600$ ∵ is a constraint, not a goal. So, it does not get any deviational variables.
- Now, we need to prioritize the goals. We do so by introducing a new cost value for each goal.
- Suppose that 1 unit of underperformance in each row costs 2,00,000 , 1,00,000, 50,000 respectively (given in question). Then we introduce the following objective function:

$$\min \quad 200k\, s_1^- + 100k\, s_2^- + 50k\, s_3^-$$

{s_i^+ would have been used if inequality was ≤}

→ This automatically forces exactly one of s_i^+ / s_i^- to be active. Hence, we do not need to concern ourselves with it separately.

- Now, we have a simple LPP, which can be solved with the simplex method.

Result: $z = 250k$, $x_1 = 6$, $x_2 = 0$, $s_1^+ = 2$, $s_3^- = 5$,
$s_2^+ = s_3^+ = s_2^- = s_1^- = 0$

☆ $s_1^- = 0 = s_2^-$ ⇒ Goals 1 & 2 are satisfied.
$s_3^- > 0$ ⇒ Goal 3 is not satisfied.

Pre-emptive Goal Programming

∃ a heirarchy of priority levels for the goals, so the goals of primary importance receive first priority attention, those of secondary importance receive second priority attention and so on.

- The relative importance of the goals are not known quantifically.
- We rank these goals in order of their importance

Goal 1 ⟶ Most important
⋮
Goal n ⟶ Least important

- We assign cost P_i to goal i such that
  $P_1 >> P_2 >> \ldots >> P_n$
  This forces goal i to be necessarily satisfied before any goal j $(j > i)$ is satisfied

- The last LP becomes:

$$\left.\begin{aligned} 7x_1 + 3x_2 + s_1^- - s_1^+ &= 40 \\ 10x_1 + 5x_2 + s_2^- - s_2^+ &= 60 \\ 5x_1 + 4x_2 + s_3^- - s_3^+ &= 35 \\ 100x_1 + 60x_2 &\leq 600 \end{aligned}\right\} \quad \begin{aligned} &x_1, x_2, s_i^+, s_i^- \geq 0 \\ &\& \\ &\min.\ P_1 s_1^- + P_2 s_2^- + P_3 s_3^- \end{aligned}$$

$\longrightarrow$ Row 0 (goal 1) min. $z_1 - P_1 s_1^- = 0$
Row 0 (goal 2) min. $z_2 - P_2 s_2^- = 0$
Row 0 (goal 3) min. $z_3 - P_3 s_3^- = 0$

Introduce decision variables in ROG* by the EROs:

$$\left\{\begin{aligned} 7x_1 + 3x_2 + s_1^- - s_1^+ &= 40 \\ 10x_1 + 5x_2 + s_2^- - s_2^+ &= 60 \\ 5x_1 + 4x_2 + s_3^- - s_3^+ &= 35 \\ 100x_1 + 60x_2 + s_4 &= 600 \end{aligned}\right.$$

$$x_1, x_2, s_i^+, s_i^- \geq 0$$

$\mathrm{ROG}i \leftarrow \mathrm{ROG}i + P_i \cdot R_i$

- Employ simplex method [with one extra check] to solve this system.
- Suppose there is a variable x_i w/ +ve (-ve) co-efficient in ROGj that needs to be minimized (maximized).
  This can be picked as the entering variable iff its co-efficient is ≥ 0 in ROG$k(<j)$ (s.t. Gk is to be minimized) & ≤ 0 in ROG$k(<j)$ (s.t. Gk is to be maximized).
- If this check fails in any goal of higher priority, then x_i cannot be picked as the entering variable.

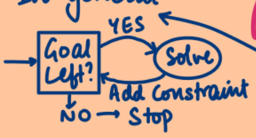
☆ Keep optimizing the i^{th} goal until there are decision variables in the goal that have a suitable co-efficient & choosing which as an entering variable will not compromise a goal of higher priority.
After goal i is done, move to goal (i+1) & so on.

☆ LINDO: Free downloadable software for solving LPPs and related problems.
How LINDO solves Goal Programming Problem:
- Stage-I: Solve for the highest priority goal.
- Stage-II: Add the solution as a constraint to the set of constraints & solve again.
- Stage-III: Add the solution to constraints & solve again for last goal.

∵ we had 3 goals

only add the goal's value as a constraint; Not the values of other variables

In general



Ex: SI: $\min z = s_1^-$ s.t. $7x_1 + 3x_2 + s_1^- - s_1^+ = 40$,
$10x_1 + 5x_2 + s_2^- - s_2^+ = 60$, $5x_1 + 4x_2 + s_3^- - s_3^+ = 35$,
$10x_1 + 60x_2 \leq 60$, all vars. ≥ 0

Solution: $z = 0 = s_1^-$

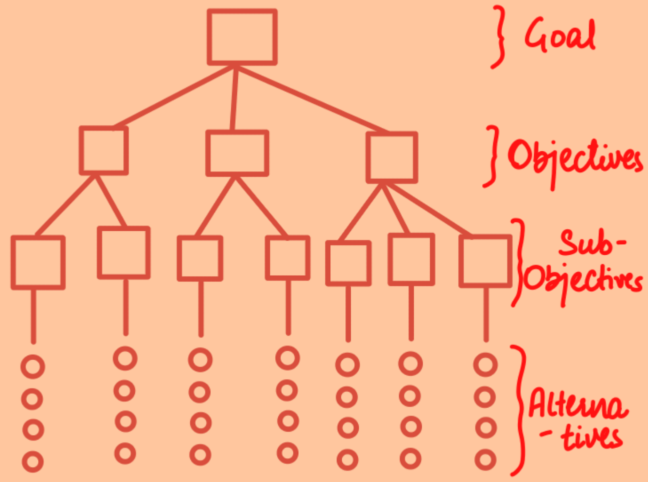
SII: $\min z = s_2^-$ s.t. $7x_1 + 3x_2 + s_1^- - s_1^+ = 40$,
$10x_1 + 5x_2 + s_2^- - s_2^+ = 60$, $5x_1 + 4x_2 + s_3^- - s_3^+ = 35$,
$10x_1 + 60x_2 \leq 60$, $s_1^- = 0$, all vars. ≥ 0

Solution: $z = 0 = s_2^-$

SIII: $\min z = s_3^-$ s.t. $7x_1 + 3x_2 + s_1^- - s_1^+ = 40$,
$10x_1 + 5x_2 + s_2^- - s_2^+ = 60$, $5x_1 + 4x_2 + s_3^- - s_3^+ = 35$,
$10x_1 + 60x_2 \leq 60$, $s_1^- = 0 = s_2^-$ all vars. ≥ 0

Solution: $z = 5 = s_3^-$, $x_1 = 6$, $x_2 = 0$, $\underbrace{s_1^- = 0 = s_2^-}_{\text{Goal 1 \& 2 were fulfilled, but Goal 3 could not be satisfied}}$.
$s_1^+ = s_2^+ = s_3^+ = 0$

Analytical Heirarchy Process (AHP)



Given a goal, some (sub)*-objectives, and some alternatives, what should we choose subject to their "scores"?

· We would like to know the "weights" (importance) of a node among its siblings with respect to its parent.
Can be found through pairwise comparisons — ratio scale priorities

Principles of AHP:
1) Decomposition ~ making the structure
2) Comparative judgements ~ finding the weights
3) Synthesis ~ evaluating the alternatives

Axioms of AHP:
1) Reciprocal axiom: If $P_c(E_A, E_B)$ is a paired comparison of elements A & B w.r.t. parent C, then $P_c(E_A, E_B) = \dfrac{1}{P_c(E_B, E_A)}$
2) Homogeneity axiom: Elements being compared should not differ too much.
3) Judgement of elements in a heirarchy does not depend on lower level elements.

Finding the weights: Suppose there are n objects and we have the following pairwise comparison matrix:

$$A = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & \cdot & & \vdots \\ \vdots & & \cdot & \\ w_n/w_1 & \cdot & \cdot & w_n/w_n \end{bmatrix}$$

and we wish to know the weights $w_1, w_2, \ldots, w_n$.

- Let $w = [w_1 \; w_2 \; \ldots \; w_n]^T$, then observe how $Aw = n\,w$. However, finding an eigenvalue-eigenvector pair is a tedious process in general.

But, observe how:

$$A[2][*] * w_1/w_2 = A[1][*] = \ldots \; A[r][*] * w_1/w_r$$

- $\therefore$ The rank of A is 1.
  $\therefore$ There is only one eigenvalue, & we already know one eigenvalue, namely n. Other eigenvalues are 0.
  Denote λ_{max} as the largest eigenvalue.
  $\therefore \quad \lambda_{max} = n$
- Our interest is to find the eigenvector corresponding to the largest eigenvalue.
  We'll work with A_{norm}

$$A_{norm} = \begin{bmatrix} \frac{w_1}{w_1+w_2+\ldots+w_n} & \frac{w_1}{sum} & \cdots & \frac{w_1}{sum} \\ \frac{w_2}{w_1+w_2+\ldots+w_n} & \frac{w_2}{sum} & \cdot & \vdots \\ \vdots & \vdots & & \cdot \\ \frac{w_n}{w_1+w_2+\ldots+w_n} & \frac{w_n}{sum} & \cdots & \frac{w_n}{sum} \end{bmatrix}$$

How to do this? Take a column, sum it, and divide each element in the column by it.

Ex: $\begin{bmatrix} w_1/w_1 \\ w_2/w_1 \\ \vdots \\ w_n/w_1 \end{bmatrix} \longrightarrow$ sum $= s = \sum w_i/w_1$

$$\frac{w_j/w_1}{s} = \frac{w_j/w_1}{\sum w_i/w_1} = \frac{w_j}{\sum w_i}$$

- The matrices in real life may not be as perfect and consistent as above. So, use A_{norm}. Find λ_{max}. Compute **consistency index** $\lambda_{max} - n / n-1$ and compare it with that of a completely random matrix. They shouldn't be close.

Transitive: $a_{ij} a_{jk} = a_{ik}$ $\forall\, i\,j\,k$

The smaller the better

☆ The entries in a row of A_{norm} may not be the same due to inconsistencies. So, we should replace each entry in a row by the row's average.
Now, take a column of A_{norm} to be w & multiply it with A. Divide Aw by w element-wise and take the average of these ratios to obtain λ_{max}.

Ex: $A = \begin{array}{c|cccc} & Sal & QL & IW & NH \\ \hline Sal & 1 & 5 & 2 & 4 \\ QL & 1/5 & 1 & 1/2 & 1/2 \\ IW & 1/2 & 2 & 1 & 2 \\ NH & 1/4 & 2 & 1/2 & 1 \end{array} \rightsquigarrow A_{norm} = \begin{bmatrix} 0.5128 & 0.5 & 0.5 & 0.533 \\ 0.1026 & 0.1 & 0.125 & 0.067 \\ 0.2564 & 0.2 & 0.25 & 0.267 \\ 0.1282 & 0.2 & 0.125 & 0.133 \end{bmatrix}$

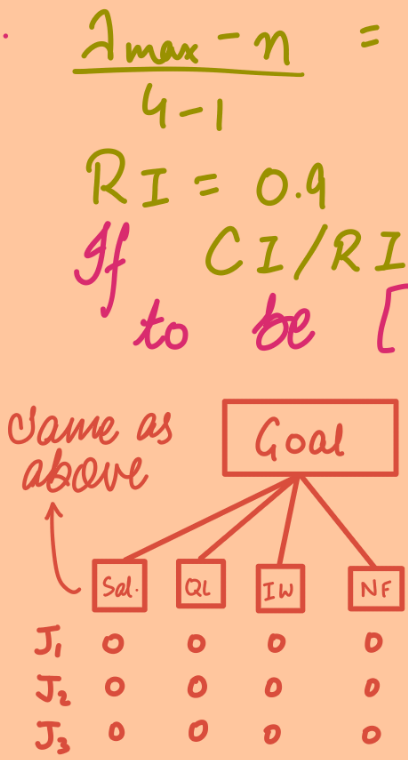
Salary, Quality of life, Interest in Work, Nearness from home

Now, A_{norm} is not perfect. So, replace each entry in a row by the row's average

$\Rightarrow w_1 = (0.5128 + 0.5 + 0.5 + 0.533)/4 = 0.5115$,
$w_2 = 0.098$, $w_3 = 0.2433$, $w_4 = 0.1466$

- Multiply the original matrix with this w
$$Aw = [2.0775 \quad 0.3159 \quad 0.9894 \quad 0.5933]$$
- Divide Aw element-wise by w & take average.
$\lambda \approx (4.061 + 4.052 + 4.0666 + 4.047)/4 = 4.05$
- $\dfrac{\lambda_{max} - n}{4-1} = \dfrac{4.05 - 4}{4-1} = 0.017 = CI$

$RI = 0.9$ (for $n = 4$) (random ratio)
- If $CI/RI < 0.1$ (consistency ratio) then it is said to be [reasonably] consistent.



Need four more matrices which describe the relative "weights"/performance of each of the jobs on each of the sub-objectives

The pairwise comparison matrix of the jobs based on salary is:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1/2 & 1 & 2 \\ 1/4 & 1/2 & 1 \end{bmatrix} \rightsquigarrow A_{norm} = \begin{bmatrix} 0.571 & 0.571 & 0.571 \\ 0.286 & 0.286 & 0.286 \\ 0.143 & 0.143 & 0.143 \end{bmatrix}$$

A_{norm} is consistent (perfectly) with J_1 salary score $= 0.571$, $J_2 = 0.286$, $J_3 = 0.143$.

Similarly, for QL: $A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 2 & 1 & 1/3 \\ 3 & 3 & 1 \end{bmatrix} \longrightarrow J_1 = 0.159, J_2 = 0.252, J_3 = 0.589$

$A_{IW} = \begin{bmatrix} 1 & 1/7 & 1/3 \\ 7 & 1 & 3 \\ 3 & 1/3 & 1 \end{bmatrix} \longrightarrow J_1 = 0.088, J_2 = 0.669, J_3 = 0.243$, $A_{NF} = \begin{bmatrix} 1 & 1/4 & 1/7 \\ 4 & 1 & 2 \\ 7 & 1/2 & 1 \end{bmatrix} \longrightarrow J_1 = 0.091, J_2 = 0.513, J_3 = 0.396$

- Score of $J_1 = \sum_{s \in \{Sal, QL, IW, NF\}} w_s \cdot w_{s,J_1} = 0.5115 \times 0.571 + 0.0986 \times 0.159 + 0.2433 \times 0.088 + 0.1466 \times 0.091$

$= 0.3425$

$J_2 = 0.5115 \times 0.286 + 0.0986 \times 0.252 + 0.2433 \times 0.669 + 0.1466 \times 0.513 = 0.4091$

$J_3 = 0.5115 \times 0.143 + 0.0986 \times 0.589 + 0.2433 \times 0.243 + 0.1466 \times 0.396 = 0.248$

$\because$ J_2 has the highest score, we select it

Data Envelopment Analysis

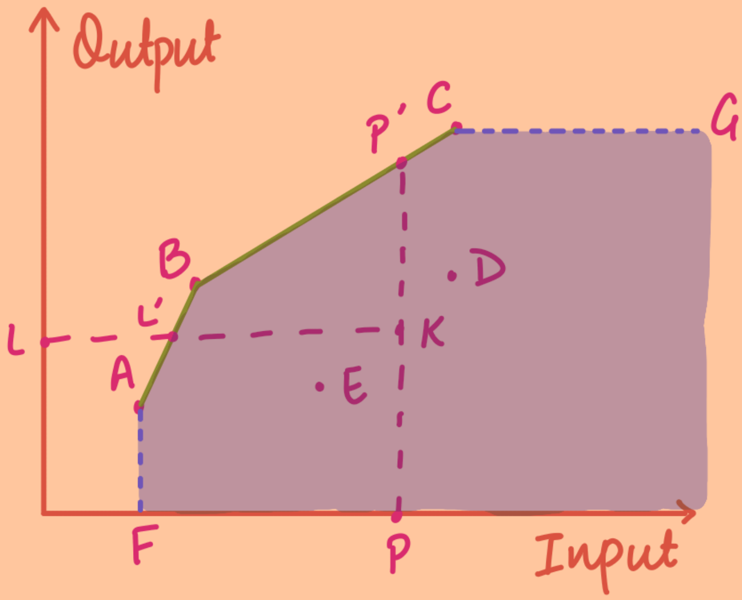
Used for comparing relative efficiencies of decision making units (DMU) which can be simple (but similar) objects like schools, factories, etc.

- The study was started by Charnes, Cooper, Rhodes and the model that so arose is known as CCR model.
- Easy to compare performance/efficiency if there is a single input and a single output but it gets complicated in a multi-input, multi-output model.

· One way to go about the latter models is parametric analysis which assumes functional form of the inputs & the outputs.
The other is non-parametric analysis under which DEA comes.
DEA works on "Production Possibility Set" (PPS).

Construction of PPS:

1.) Interpolation between feasible input-output correspondences leads to new input-output correspondences which are feasible in principle.
2.) Inefficient production is possible.
3.) PPS is the smallest set meeting the above two assumptions & containing all input-output correspondences observed at the units being assessed.



Suppose we are given DMUs' input-output pairs A, B, C, D, E
· We envelope these points using the two assumptions above (A1 & A2) and call the shaded region the PPS

· A1: All points lying on the line joining the given points are in PPS.

· A2: All lesser efficient points (towards right & bottom) are in PPS.

Use more input

produce less output

- All points lying on the lines AB, BC are efficient. All other points are inefficient.

- In general, there may be other curves used for interpolation, but assuming CRS (constant returns to scale), a straight line is the one to be used.

☆ Efficiency of K: Input Efficiency = LL'/LK
Output Efficiency = PP'/PK

Pareto Efficiency

- Input Orientation: A DMU is Pareto efficient if it is not possible to lower any one of its input levels without increasing atleast another one of its input levels and/or without lowering at least one of its output levels.
- Output Orientation: A DMU is Pareto efficient if it is not possible to raise any one of its output levels without lowering at least another one of its output levels and/or without increasing atleast one of its input levels.

Technical Efficiency

- Input Efficiency: Let $(\bar{x}, \bar{y})$ be an input-output correspondence.
Here, $\bar{x}$ is the vector of inputs & $\bar{y}$ is

the vector of outputs.

$$L(\bar{y}) := \{ x \mid x \text{ can produce } \bar{y} \}$$

$$\bar{x} \in L(\bar{y}) \Rightarrow L(\bar{y}) \neq \phi$$

$$\theta^* = \max_{\theta} \{ \theta \mid x/\theta \in L(\bar{y}),\ 0 < \theta \in \mathbb{R} \}$$

Technical input efficiency of the DMU with input-output correspondence $(\bar{x}, \bar{y})$ is defined as $\dfrac{\bar{x}/\theta^*}{\bar{x}} = \boxed{\dfrac{1}{\theta^*}}$

- Output Efficiency:

$(\bar{x}, \bar{y})$ same as above.

$$L(\bar{x}) := \{ y \mid y \text{ can be produced using } \bar{x} \}$$

$$\bar{y} \in L(\bar{x}) \Rightarrow L(\bar{x}) \neq \phi$$

$$\theta^* = \max_{\theta} \{ \theta \mid \theta y \in L(\bar{x}),\ 0 < \theta \in \mathbb{R} \}$$

Technical output efficiency of DMU $(\bar{x}, \bar{y})$ is: $\dfrac{\bar{y}}{\theta^* \bar{y}} = \boxed{\dfrac{1}{\theta^*}}$

Note: Technical input & output efficiencies are weaker conditions compared to Pareto Efficiency (input or output).

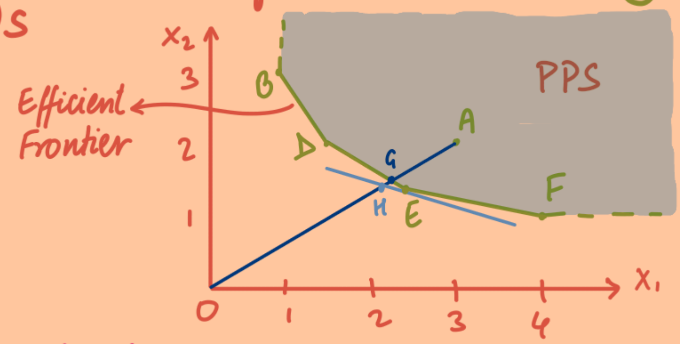
☆ Technical Efficiency does not take into account the "relative" importance of a specific input (or output) dimension. Allocative Efficiency does so.

Allocative Efficiency

It incorporates the price / cost of the output / input parameters.

Example: Suppose there are two inputs (x_1, x_2) to produce one unit of output (y)

A, B, D, E, F are DMUs (Output = 1) given to us.



- All points on BDEF are Pareto-Efficient.
- All points on the efficient frontier have technical efficiency = 1; Even the points above B & towards the right of F are also technically efficient but not Pareto Efficienct, hence the note.
- To find Technical Efficiency of DMU A, we find the intersection point G of OA & Efficient Frontier. Then, efficiency is OG/OA.

Now, suppose 1 unit of input x_1 costs 4, and x_2 10. Then, the iso-cost line is given by $4x_1 + 10x_2 = k$

To find allocative efficiency (a.k.a. price efficiency) of A, we find the tangent to efficient frontier along the isocost lines.

Then, find the intersection point of the tangent & OA. Call it H.

The allocative efficiency of A is OH/OG.

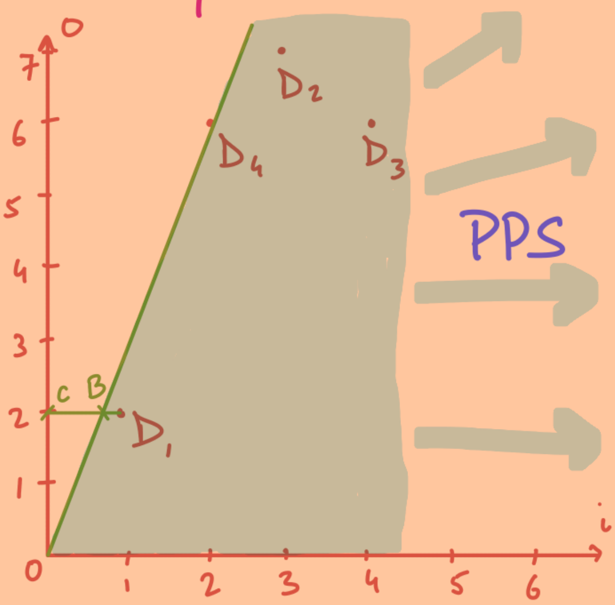
☆ The Overall efficiency is simply the product of Technical Efficiency and Allocative Efficiency, i.e., $\frac{OG}{OA} \cdot \frac{OH}{OG} = \frac{OH}{OA}$

DEA under CRS

CRS → Constant Returns to Scale

Basic principles & construction of PPS:

1) Interpolation between feasible input-output correspondences leads to input-output correspondences feasible in principle.
2) Inefficient production is possible.
3) The transformation of inputs to outputs is characterised by CRS, i.e., if $(\bar{x}, \bar{y})$ is a feasible input-output correspondence then for $\alpha > 0$, $(\alpha\bar{x}, \alpha\bar{y})$ is also a feasible input-output correspondence.
4) No output is possible unless some input is used. (No Free Lunch)
5) The PPS is the smallest set satisfying these assumptions & containing all observed input-output correspondences.



Ex: Consider the case of 4 DMUs which use a single input to secure a single output: $D_1(1, 2)$, $D_2(3, 7)$, $D_3(4, 6)$, $D_4(2, 6)$

- Compute input efficiency of DMU D_1.

  Fix output & draw horizontal line: CB/CD.

- The PPS can be represented using LP:

$$\left\{ \begin{array}{l} (x, y) : \quad x \geq 1\cdot\lambda_1 + 3\cdot\lambda_2 + 4\cdot\lambda_3 + 2\cdot\lambda_4, \\ \lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0, \; y \leq 2\cdot\lambda_1 + 7\cdot\lambda_2 + 6\cdot\lambda_3 + 6\cdot\lambda_4 \end{array} \right\}$$

· Moreover, LPP can be used to measure technical input efficiency of DMU D_2:

$$\min. \ z, \ \text{s.t.} \ 1\cdot\lambda_1 + 3\cdot\lambda_2 + 4\lambda_3 + 2\lambda_4 \le 3z$$
$$2\cdot\lambda_1 + 7\cdot\lambda_2 + 6\lambda_3 + 6\lambda_4 \ge 7$$
$$\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ge 0, \quad z: \text{free variable}$$

Then, Input Efficiency will be $z^* (\le 1)$

Duality of an LPP

· Normal max problem $\xleftrightarrow{\text{Dual}}$ Normal min problem

All constraints $\le$ | All constraints $\ge$

$\max. \ z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$	$\min. \ w = b_1y_1 + b_2y_2 + \ldots + b_my_m$
s.t. $a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1$	s.t. $a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \ge c_1$
$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2$	$a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \ge c_2$
$\vdots$	$\vdots$
$a_{m1}x_1 + \ldots + a_{mn}x_n \le b_m$	$a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_n \ge c_n$
$x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0$	$y_1 \ge 0, y_2 \ge 0, \ldots, y_n \ge 0$

☆ Any non-normal max. problem can be converted to a normal max. problem:

$\ge$: Multiply constraint w/ -1; $=$: Split to $\le, \ge$

u.r.s. : Use $x := x_1 - x_2$, $x_1 \ge 0$, $x_2 \ge 0$

Similarly, for any non-normal min. problem.

· The original problem is called the primal problem and the newly formed, other, LPP is called the dual.

· One way to find the dual of a problem.

	$x_1 \ge 0$	$x_2 \ge 0$	$\ldots$	$x_n \ge 0$	max. z
$y_1 \ge 0$	a_{11}	a_{12}	$\ldots$	a_{1n}	$(\le) b_1$
$y_2 \ge 0$	a_{21}	a_{22}	$\ldots$	a_{2n}	$(\le) b_2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$y_m \ge 0$	a_{m1}	a_{m2}	$\ldots$	a_{mn}	$(\le) b_m$
min. w	$(\ge) c_1$	$(\ge) c_2$	$\ldots$	$(\ge c_n)$	

Read Down → Get normal min LPP

Dual

Read Across → Get normal max LPP

- Dual of a dual is the primal itself.

Weak Duality (Theorem)

Let max z & min w be two dual LPPs, and x, y be feasible solutions to the LPPs respectively. Then, we always have:

$$\boxed{z|_x \leq w|_y}$$

Proof: $\sum_j a_{ij} x_j \leq b_i \Rightarrow \sum_j y_i a_{ij} x_j \leq b_i y_i \rightarrow$ Sum over i

$z|_x = \sum_j c_j x_j \leq \sum_j \sum_i y_i a_{ij} x_j \leq \sum_i b_i y_i = w|_y$

$c_j \leq \sum_i y_i a_{ij} \Rightarrow c_j x_j \leq \sum_i y_i a_{ij} x_j \rightarrow$ Sum over j

- Most of the times, $\exists$ interpretations of the dual in the same context.

☆ If $\exists$ feasible y_0 of dual w (min.), then we can bound the value of primal, i.e., $z \leq w|_{y_0}$ $\forall x$

☆ If $\exists$ feasible x_0 of primal z (max.), then we can bound the value of dual, i.e., $z|_{x_0} \leq w$ $\forall y$

☆ Let there be feasible solutions $\bar{x}$, $\bar{y}$ s.t. $c\bar{x} = \bar{y}b$, then $\bar{x}$, $\bar{y}$ are optimal solutions.

Dual Theorem

- Primal unbounded $\Rightarrow$ Dual infeasible

  Dual unbounded $\Rightarrow$ Primal infeasible

☆ Optimal table of primal LPP can give insight to dual optimal values:

$$i^{th} \text{ Constraint} \begin{cases} \leq & \text{— + Co-eff. of } s_i \\ \geq & \text{— − Co-eff. of } e_i \\ = & \text{— } M + \text{Co-eff. of } a_i \end{cases} \text{ in Row 0} = y_i$$

· Similarly, the technical output efficiency of DMU D_1 can also be found using LPP:

max. z s.t. $\lambda_1 + 3\lambda_2 + 4\lambda_3 + 2\lambda_4 \le 1$,

$2\lambda_1 + 7\lambda_2 + 6\lambda_3 + 6\lambda_4 \ge 2z$,

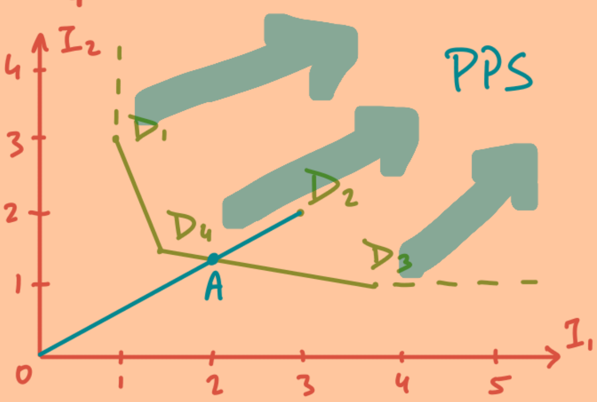
$\lambda_i \ge 0$ decision vars., z free variable.

$z^* \ge 1$; Efficiency is : $1/z^*$

☆ Under CRS, Technical Input Efficiency is equal to Technical Output Efficiency.

Ex: Single-output-multi-input under CRS. Measure TIE.
They produce $\underbrace{1 \text{ unit}}$ of output.
If they do not, we can scale them by dividing by output.

DMU	Input 1	Input 2
D_1	1	3
D_2	3	2
D_3	3.75	1
D_4	1.5	1.5



· All DMUs D_1, D_3, D_4 have TIE efficiency 1 (and are also pareto-efficient).

· For DMU D_2 : $\dfrac{OA}{OD_2} = \dfrac{11}{16}$

· For any point (x_1, x_2, y) in PPS, we have:

$x_1 \ge \lambda_1 \cdot 1 + \lambda_2 \cdot 3 + \lambda_3 \cdot 3.75 + \lambda_4 \cdot 1.5$, $y \le \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$,

$x_2 \ge \lambda_1 \cdot 3 + \lambda_2 \cdot 2 + \lambda_3 \cdot 1 + \lambda_4 \cdot 1.5$, $\lambda_i \ge 0$

· To find technical input efficiency of DMU D_2 using LPP:

min. z s.t. $\lambda_1 + 3\lambda_2 + 3.75\lambda_3 + 1.5\lambda_4 \le 3z$,

$3\lambda_1 + 2\lambda_2 + \lambda_3 + 1.5\lambda_4 \leq 2z$. $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \geq 1$

Adding slack variables :

$$\begin{cases} \lambda_1 + 3\lambda_2 + 3.75\lambda_3 + 1.5\lambda_4 + s_1 = 3z \\ 3\lambda_1 + 2\lambda_2 + \lambda_3 + 1.5\lambda_4 + s_2 = 2z \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - s_3 = 1 \end{cases}$$

Rearranging :

$$\lambda_1 + 3\lambda_2 + 3.75\lambda_3 + 1.5\lambda_4 = 3z - s_1$$
$$3\lambda_1 + 2\lambda_2 + \lambda_3 + 1.5\lambda_4 = 2z - s_2$$
$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 + s_3$$

The LHS represents true-input level / true output level of a point.

- Now, we would like to minimize the following quantity: $\min \; z - \epsilon(s_1 + s_2 + s_3)$

infinitesimally small

Minimize (ratio) input used with first and foremost priority. With second priority, we would like to min. $-s_1, -s_2$ and max. s_3, i.e., min. $-(s_1 + s_2 + s_3)$ in order to dec. input / inc. output. Essentially, $z \sim$ Scale of I/O used (1st priority), $s_i \sim$ Amount of I/O used (2nd priority)

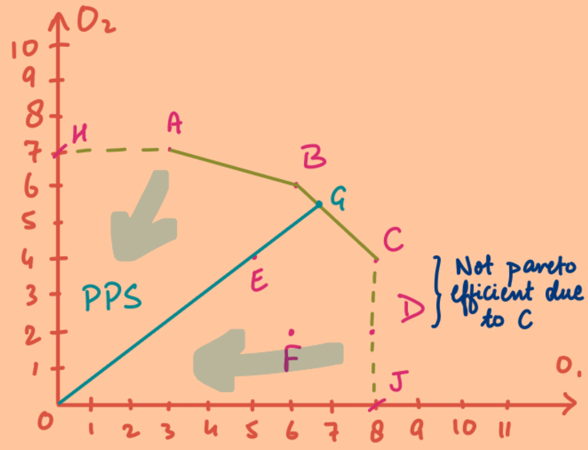
☆ If $z^* < 1$, then the DMU is neither technically efficient nor pareto-efficient, but z^* gives the technical efficiency.

If $z^* = 1$, then the DMU is technically efficient. Even if any one of $s_i \neq 0$, then the point is not pareto efficient.

Otherwise, if $z^* = 1$ & all $s_i = 0$, then the point is pareto efficient as well.

Ex: Single-Input-Multiple-Output under CRS. Measure TOE

DMU	Input (1000's)	O_1	O_2
A	$1 \leftarrow 300$	$(3,7) \leftarrow 900$	2100
B	$1 \leftarrow 250$	$(6,6) \leftarrow 1500$	1500
C	$1 \leftarrow 320$	$(8,4) \leftarrow 2560$	1280
D	$1 \leftarrow 400$	$(8,2) \leftarrow 3200$	800
E	$1 \leftarrow 100$	$(5,4) \leftarrow 500$	400
F	$1 \leftarrow 80$	$(6,2) \leftarrow 480$	160



- Pareto Efficient: $\{A-B,\ B-C\}$
- Technically efficient: $\{H-A,\ A-B,\ B-C,\ C-D-J\}$
- Efficiency of E graphically: $\dfrac{OE}{OG}$
- Efficiency of E using LPP:

$$300\lambda_1 + 250\lambda_2 + 320\lambda_3 + 400\lambda_4 + 100\lambda_5 + 80\lambda_6 \leq 100$$
$$900\lambda_1 + 1500\lambda_2 + 2560\lambda_3 + 3200\lambda_4 + 500\lambda_5 + 480\lambda_6 \geq 500h$$
$$200\lambda_1 + 1500\lambda_2 + 1280\lambda_3 + 800\lambda_4 + 400\lambda_5 + 160\lambda_6 \geq 400h$$

$\lambda_i \geq 0$, h free. Add $s_1, -s_2, -s_3$ to LHS and re-arrange to get true input/output on LHS. Then, we would like to maximize the scale (h) of output first and then $(s_1 + s_2 + s_3)$. Thus, we maximize $h + \underbrace{\epsilon}_{\text{infinitesimal}}(s_1 + s_2 + s_3)$

★ $1/h^*$ gives T.O.E. of DMU E.

If $1/h^* = 1$, then its T.O.E. is 1.

If all $s_i = 0$ & $h^* = 1$ then it is Pareto efficient. But, even if a single $s_i \neq 0$ then T.O.E. remains 1, but it won't be Pareto efficient.

If $1/h^* < 1$, then it isn't technically efficient

Multi-Input-Multi-Output (TIE under CRS)

Suppose there are [n] DMUs using m inputs to secure s outputs.

Let x_{ij} & y_{rj} be the level of the i^{th} input & r^{th} output respectively of j^{th} DMU.

Suppose we want to find TIE of DMU j.

Construct the following table:

	DMU	1	2	3	… i (Input) …	m	1	2	3	… r (Output) …	s
λ_1	1	x_{11}	x_{21}	x_{31}	… x_{i1} …	x_{m1}	y_{11}	y_{21}	y_{31}	… y_{r1} …	y_{s1}
λ_2	2	x_{12}	x_{22}	x_{32}	⋮	⋮	y_{12}	y_{22}	⋮	⋮	⋮
⋮	⋮	⋮					⋮				
λ_n		x_{1n}	…	…	…	x_{mn}	y_{1n}	…	…	…	y_{sn}

↓

Infinitesimally small

LPP → $\min.\ z - \epsilon(s_1^- + s_2^- + \ldots + s_m^- + s_1^+ + \ldots + s_s^+)$

$(s_1^- +)\quad \lambda_1 x_{11} + \lambda_2 x_{12} + \ldots + \lambda_n x_{1n} \leq z x_{1j}$,

$(s_2^- +)\quad \lambda_1 x_{21} + \lambda_2 x_{22} + \ldots + \lambda_n x_{2n} \leq z x_{2j}$,

$\vdots$

$(s_m^- +)\quad \lambda_1 x_{m1} + \lambda_2 x_{m2} + \ldots + \lambda_n x_{mn} \leq z x_{mj}$,

$(-s_1^+)\quad \lambda_1 y_{11} + \lambda_2 y_{12} + \ldots + \lambda_n y_{1n} \geq y_{1j}$,

$\vdots$

$(-s_s^+)\quad \lambda_1 y_{s1} + \lambda_2 y_{s2} + \ldots + \lambda_n y_{sn} \geq y_{sj}$,

$\lambda_1, \lambda_2 \ldots \lambda_n \geq 0$, z u.r.s.

Contracting factor

z being u.r.s. will be very significant in value-based DEA

It will come in [0,1]. We're not constraining z further ∵ solution guarantees that.

- At optimality, if $z = 1$ & $s_i^- = 0\ \forall i$, $s_r^+ = 0\ \forall r$, then TIE of DMU j is 1 & it is Pareto-efficient.
- If $z = 1$, but not all $s_i^{+/-} = 0$, i.e., at least one $s_i \neq 0$ or $s_r \neq 0$ then its TIE is 1 but DMU is not Pareto-efficient.
- If $z < 1$, then it is not Pareto-efficient & z gives the TIE of DMU j.

- For TOE, use h, the expansion factor, at appropriate places. TOE will be given by $1/h$.

Ex. A mail order company operates 5 regional distribution centres. Last complete quarter:

Centre	Labour hrs. (1000s)	Capital employed (millions)	Delivery distance (1000 kms.)	Package delivered (1000s)
1	4.1	2.3	4.3	90
2	3.8	2.4	3.9	102
3	4.4	2.0	4.1	96
4	3.2	1.8	5.2	110
5	3.4	3.4	4.2	120
	input	input	output	output

Find TIE & TOE of Centre 1

TIE: $z^* \longleftarrow \min \; z - \epsilon(s_1^- + s_2^- + s_1^+ + s_2^+)$

$(s_1^-+)\quad 4.1\lambda_1 + 3.8\lambda_2 + 4.4\lambda_3 + 3.2\lambda_4 + 3.4\lambda_5 \leq 4.1z$

$(s_2^-+)\quad 2.3\lambda_1 + 2.4\lambda_2 + 2.0\lambda_3 + 1.8\lambda_4 + 3.4\lambda_5 \leq 2.3z$

$(-s_1^+)\quad 4.3\lambda_1 + 3.9\lambda_2 + 4.1\lambda_3 + 5.2\lambda_4 + 4.2\lambda_5 \geq 4.3$

$(-s_2^+)\quad 90\lambda_1 + 102\lambda_2 + 96\lambda_3 + 110\lambda_4 + 120\lambda_5 \geq 90$

TOE: $1/h^* \longleftarrow \max \; h + \epsilon(s_1^- + s_2^- + s_1^+ + s_2^+)$

$(s_1^-+)\quad 4.1\lambda_1 + 3.8\lambda_2 + 4.4\lambda_3 + 3.2\lambda_4 + 3.4\lambda_5 \leq 4.1$

$(s_2^-+)\quad 2.3\lambda_1 + 2.4\lambda_2 + 2.0\lambda_3 + 1.8\lambda_4 + 3.4\lambda_5 \leq 2.3$

$(-s_1^+)\quad 4.3\lambda_1 + 3.9\lambda_2 + 4.1\lambda_3 + 5.2\lambda_4 + 4.2\lambda_5 \geq 4.3h$

$(-s_2^+)\quad 90\lambda_1 + 102\lambda_2 + 96\lambda_3 + 110\lambda_4 + 120\lambda_5 \geq 90h$

Value Based DEA Model

This is essentially the Dual of the envelopment model we make using the rules above.

- Measurement of efficiency in terms of associated values. Shall see how to interpret the Dual to get value based DEA model with "imputed values".

- Recall that the TIE of DMU j was computed by solving the following LPP:

$$\lambda_1 x_{11} + \lambda_2 x_{12} + \ldots + \lambda_n x_{1n} + s_1^- = k\, x_{1j}$$

$$\begin{array}{rl|l}
 & & \text{For Dual} \\
\Rightarrow & k\, x_{1j} - \lambda_1 x_{11} - \lambda_2 x_{12} \ldots - \lambda_n x_{1n} - s_1^- = 0 & v_1 \\
 & k\, x_{2j} - \lambda_1 x_{21} - \ldots - \lambda_n x_{2n} - s_2^- = 0 & v_2 \\
 & \vdots & \vdots \\
 & k\, x_{mj} - \lambda_1 x_{m1} - \ldots - \lambda_n x_{mn} - s_m^- = 0 & v_m \\
 & \lambda_1 y_{11} + \lambda_2 y_{12} + \ldots + \lambda_n y_{1n} - s_1^+ = y_{1j} & u_1 \\
 & \vdots & \vdots \\
 & \lambda_1 y_{s1} + \lambda_2 y_{s2} + \ldots + \lambda_n y_{sn} - s_s^+ = y_{sj} & u_s
\end{array}$$

& $\min.\ k - \epsilon(s_1^- + s_2^- + \ldots + s_m^- + s_1^+ \ldots + s_s^+)$

Leads to equality $\leftarrow$ $k : \text{u.r.s.}$, $\quad \lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$ $\rightarrow$ Leads to inequality

$$\max.\ 0\cdot v_1 + 0\cdot v_2 + \ldots + 0\cdot v_m + y_{1j}\cdot u_1 + y_{2j}\cdot u_2 + \ldots + y_{sj}\cdot u_s$$

$$\begin{array}{ll}
\text{s.t.} & x_{1j}\cdot v_1 + x_{2j}\cdot v_2 + \ldots + x_{m1}\cdot v_m = 1 \\
 & -x_{11}\cdot v_1 - x_{21}\cdot v_2 - \ldots - x_{m1}\cdot v_m + y_{11}\cdot u_1 + y_{21}\cdot u_2 + \ldots + y_{s1}\cdot u_s \leq 0 \\
 & -x_{12}\cdot v_1 - x_{22}\, v_2 - \ldots - x_{m2}\cdot v_m + y_{12}\cdot u_1 + y_{22}\cdot u_2 + \ldots + y_{s2}\cdot u_s \leq 0 \\
 & \vdots \\
 & -x_{1n}\cdot v_1 - x_{2n}\, v_2 - \ldots - x_{mn}\cdot v_m + y_{1n}\cdot u_1 + y_{2n}\cdot u_2 + \ldots + y_{sn}\cdot u_s \leq 0 \\
 & -v_1 \leq -\epsilon, \quad -v_2 \leq -\epsilon, \quad \ldots, \quad -v_m \leq -\epsilon, \\
 & \quad -u_1 \leq -\epsilon, \quad -u_2 \leq -\epsilon, \quad \ldots, \quad -u_s \leq -\epsilon
\end{array}$$

- How the dual is made:
  - Make a variable corresponding to each constraint (u_i, v_r)
  - Obj. func. min $\rightarrow$ max. & multiply the newly formed vars. with the constants on their equation's RHS. Add all these terms to get the obj. function
  - Then, for each var, multiply v_i / u_r with the var's co-efficient in v_i's/u_r's constraint. Add all these. If the var. was $\leq$ then

the new constraint shall be $\geq$. Similarly, $\leq$ if it was $\geq$ in the primal and $=$ if the variable was u.r.s. in the primal.

The constant on the R.H.S. of the new constraint shall be the variable's co-efficient in the objective function. For example:

k, min. $1 \cdot k - \epsilon(\cdot)$, k u.r.s. $\rightsquigarrow \sum x v = 1$

λ, min. $0 \cdot \lambda + (\cdot)$, $\lambda \geq 0 \rightsquigarrow -\sum x v + \sum y u \leq 0$

s, min. $\epsilon s - (\cdot)$, $s \geq 0 \rightsquigarrow -v \leq -\epsilon, -u \leq -\epsilon$

- Each constraint of the dual has an interpretation:
  - For any DMU, $\sum v x$ is the input, and $\sum u y$ is the output.
  - A general measure of efficiency is $\frac{\text{OUTPUT}}{\text{INPUT}}$. The input of DMU of interest is normalized to 1. Thus, in the optimal function we are essentially maximizing the output.

    If it had not been normalized, we would be maximizing $\frac{?}{0}$, which would be an NLP (non-linear programming) problem.
  - For all other DMU's, we get a constraint of the kind $-\sum v x + \sum u y \leq 0$ ~ output $\leq$ input
  - Constraints of the kind $-v \leq -\epsilon$, $-u \leq -\epsilon$ are essentially $v > 0$, $u > 0$, i.e., there is no free input / output.
- $\because$ Optimal value of a primal LPP and dual LPP is the same by duality theorem, k can be approximated (not exactly due to ϵ) by the optimal value of the dual.
- Suppose the primal is min $k - \epsilon(\cdot)$ and the optimal value of dual is exactly 1.

Then, $\because k \leq 1$, we have that $k=1$, i.e., the DMU is technically efficient. But, we also have that $\epsilon(\cdot) = 0 \Rightarrow s_i^{+/-} = 0 \ \forall i$, so that the DMU is Pareto efficient as well.

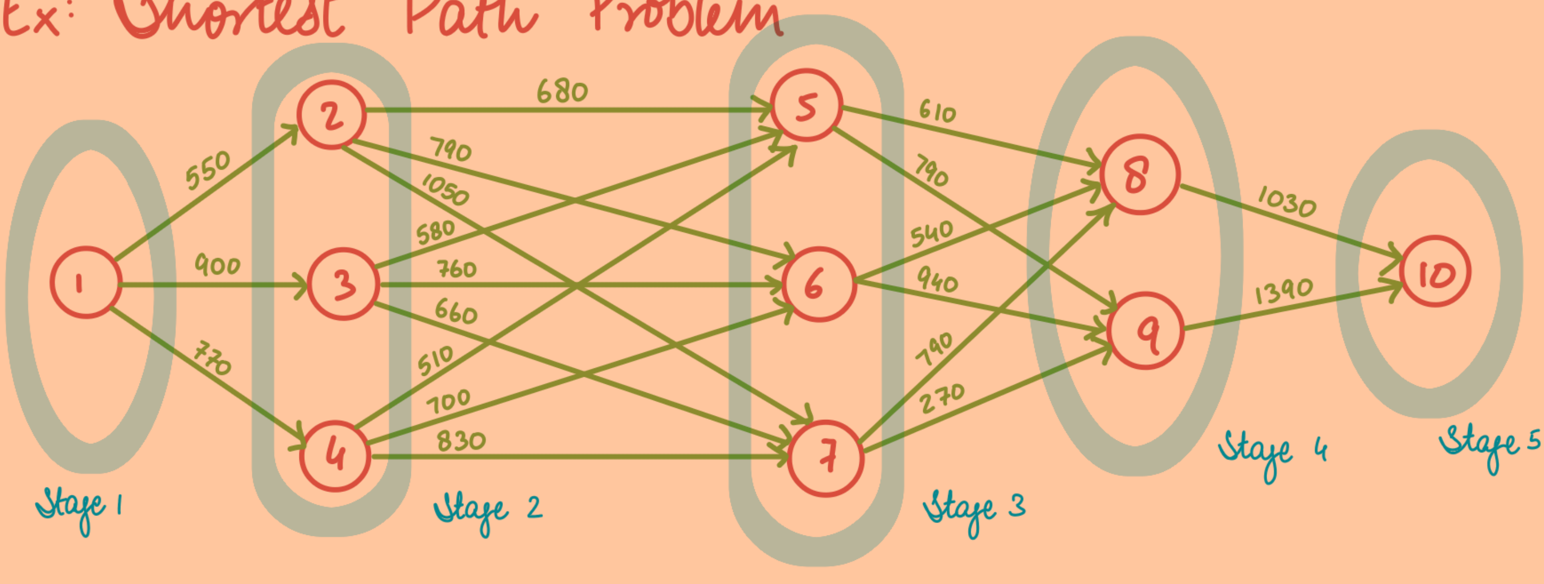
Dynamic Programming

Large Problem is divided into a number of subproblems.

- Deterministic Dynamic Programming
  Parameters are known with certainty.
- Probabilistic Dynamic Programming
  (Some) Parameters have some probability assigned to them.
- Markov Dynamic Programming
  The "timeframe" of decisions is infinite / Number of decision stages is unlimited / Long term decision making.

- Most of the times, "backward induction" is used to solve DP problems.
  "Forward Induction" is also used sometimes.

Ex: Shortest Path Problem



c_{ij} : dist. from city i to city j [given]
Find the shortest path from 1 to 10

- We begin by classifying the cities into stages. $\{1\}$ = Stage 1. Similarly, $\{2, 3, 4\}$, $\{5, 6, 7\}$, $\{8, 9\}$, $\{10\}$. Further, let $f_t(i)$ denote the length of the shortest path from city i to city 10, given that city i is in stage t.
- We only need to make decisions in stages 1, 2, 3, 4.

Stage 4: Determine shortest path from 8, 9 to 10.
$\because$ there is only one path from 8 to 10, it becomes the shortest path. Hence, we have
$f_4(8) = c_{8,10} = 1030$ & optimal decision: 8 — 10
Illy, $f_4(9) = c_{9,10} = 1390$ & optimal decision: 9 — 10

Stage 3: We work one stage backwards, and find the shortest path from each of the cities of S_3 to city 10.
Note that, for eg. from city 5 to city 10 can either be through the path 5 — 8 — 10 or 5 — 9 — 10. Hence,

$$f_3(5) = \min.\begin{cases} c_{58} + f_4(8) = 1640 \\ c_{59} + f_4(9) = 2180 \end{cases} \Rightarrow f_3(5) = 1640 \text{ \& optimal decision } 5-8$$

Similarly, $f_3(6) = \min.\{c_{68} + f_4(8), c_{69} + f_4(9)\} = 1570$ & optimal decision 6 — 8, and $f_3(7) = 1660$ w/ optimal decision 7 — 9.

Stage 2: In a similar way as above,
$f_2(2) = \min\{c_{25} + f_3(5), c_{26} + f_3(6), c_{27} + f_3(7)\} = 2320$, optimal decision: 2 — 5; $f_2(3) = 2220$ &

optimal decision: 3 — 5 ; $f_2(4) = 2150$ w/ 4 — 5.

Stage 1: $f_1(1) = 2870$ w/ 1 — 2

- With all the stages' computations complete, the optimal path is traced as follows:
  - → Go from 1 to 2.
  - → Refer to S_2 calculations: Go from 2 to 5.
  - → Refer to S_3 calculations: Go from 5 to 8.
  - → Refer to S_4 calculations: Go from 8 to 10.
- Thus, the optimal path is 1 — 2 — 5 — 8 — 10 and the length of the path is 2870.

☆ Works because of Bellman's Principle of Optimality:

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Characteristics of Dynamic Programming

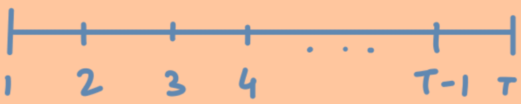
1) There are multiple stages.
   A decision has to be made in a stage.
2) Each stage has a number of states.
   This is required to make a decision in the particular stage.
3) A decision in a stage at a particular state transitions it into a state in the next stage.

4) Principle of Optimality: Decision in the current stage should be independent of the previously reached state.

5) Recursion: The cost or the reward at stages t, $t+1$, ... is related to the cost / reward at stages $t+1$, $t+2$, ...

☆ Stages can be continuous or discrete

Ex 1: Inventory Problem



The entire decision making period is divided into some timeframes. Demand during each period is known at the beginning of period 1. Limitations on capacity of inventory. Can overproduce. Must decide how much to produce at the beginning of each period.

Each period's demand must be met from inventory or current production.

No inventory at the beginning of stage 1.

When producing, a fixed cost and variable production cost is incurred; similarly storing a unit has some cost.

Goal is to minimize the total cost of meeting demands for periods $1 \dots T$ on time

A concrete example: Set-up cost = 3, Variable cost = 1, Holding cost = 0.5, Demand -

Month 1 = 1, M2 = 3, M3 = 2, M4 = 4

Formulate the DP as follows:

Stage: The month we are in.

State: The number of items in the inventory.

Decision: How much to produce

$f_t(i)$:= min. cost of prod. to fulfill demand of months $t, t+1, \dots, 4$ in stage i (there are i units in the inventory)

$c(x) = \begin{cases} 0, & x=0 \\ 3+x, & x>0 \end{cases}$:= cost of producing x units during a period.

$x_t(i)$:= No. of units to be produced during period t if i units of inventory are on hand.

Stage 4: In the beginning of each stage, the state can be $0, 1, 2, 3, 4$; $M_4 = 4$.

$\because$ this is the last state, produce just enough to meet this month's demand — no need to overproduce $\Rightarrow i + x < 5$

	0	1	2	3	4
$4-i \equiv x_4(i)$	4	3	2	1	0
$c(x_4(i)) \equiv f_4(i)$	$7=3+4$	6	5	$4=3+1$	0

Stage 3: $M_3 = 2$. Let x denote current month's production $\longrightarrow$ $x + i \geq M_3 \Rightarrow 2 \leq i+x$.

Also, $i+x-2$ is the amount of left over inventory, capacity is $4 \Rightarrow i+x-2 \leq 4 \Rightarrow i+x \leq 6$

Inventory cost is $(i+x-2)/2$, and production cost is $c(x)$.

$$f_3(i) = \min_{\substack{x \\ 2 \leq i+x \leq 6}} \{ (i+x-2)/2 + c(x) + f_4(i+x-2) \}$$

Ex: $f_3(1) \longrightarrow 1 \le x \le 5 = 6 - i = 6 - 1$

$$f_3(1) = \min \{ 0+4+7 |_{x=1},\ 1/2+5+6 |_{x=2},\ 1+6+5,\ 3/2+7+4,\ \underline{2+8+0} \} = 10$$

$f_3(1) = 10$ at $x_3(1) = 5$

Similarly,

$f_3(0) = 12$ at $x_3(0) = 2$
$f_3(2) = 7$ at $x_3(2) = 0$
$f_3(3) = 13/2$ at $x_3(3) = 0$
$f_3(4) = 6$ at $x_3(4) = 0$

Stage 2: $M_2 = 3 \Rightarrow i + x \geq 3$

Capacity of inventory $= 4 \Rightarrow i + x - 3 \le 4$

Computing $f_2(i)$:

$f_2(0) = 16$ at $x_2(0) = 5$
$f_2(1) = 15$ at $x_2(1) = 4$
$f_2(2) = 14$ at $x_2(2) = 3$
$f_2(3) = 12$ at $x_2(3) = 0$
$f_2(4) = 21/2$ at $x_2(4) = 0$

Stage 1: $M_1 = 1 \Rightarrow i + x \geq 1$ & $i + x - 1 \le 4$.

No inventory at beginning of $M_1 \Rightarrow i = 0$

$f_1(i) \longrightarrow f_1(0) = 20$ at $x_1(0) = 1$

Optimal Decisions: Minimum Cost is 20.

At stage 1, produce 1 unit(s) & $M_1 = 1$
$\Rightarrow$ 0 in inventory for next stage
At stage 2, produce 5 unit(s) & $M_2 = 3$
$\Rightarrow$ 2 in inventory for next stage
At stage 3, produce 0 unit(s) & $M_3 = 2$
$\Rightarrow$ 0 in inventory for next stage

At stage 4, produce 4 unit(s).

$\therefore$ We produce (1, 5, 0, 4) at Total Cost 20

Ex 2: Equipment Replacement Problem

A shop always needs to have an engine analyzer available.

A new engine costs 1000

The cost m_i of maintaining an engine in its i^{th} year of operation is:

$m_1 = 60, \quad m_2 = 80, \quad m_3 = 120$

An analyzer may be kept for upto 3 years

After i years of use, it may be traded in for a new one for a salvage value s_i as: $\quad s_1 = 800, \quad s_2 = 600, \quad s_3 = 500$

Given that a new machine must be purchased now, the shop wants to determine a policy that minimizes net costs for the next 5 years.

Time: 0 1 2 3 4 5

Formulate the type as follows:

Stages: The different years or time periods

States: Age of the engine

Decisions: Sell (trade) or Keep (maintain)

$f_t(x)$:= Minimum cost incurred from time t to time 5 given that at time t, the shop has an x-year old analyzer

The problem is over at $t=5$, so we sell the machine $\Rightarrow f_5(i) = -s_i$

$\Rightarrow f_5(1) = -800, \quad f_5(2) = -600, \quad f_5(3) = -500$

For $t = 1, 2, 3, 4$ we have:

$$f_t(3) = -500 + 1000 + 60 + f_{t+1}(1). \qquad \text{Sell}$$

$$f_t(2) = \min. \begin{cases} -600 + 1000 + 60 + f_{t+1}(1). & \text{Sell} \\ 120 + f_{t+1}(3). & \text{Keep} \end{cases}$$

$$f_t(1) = \min. \begin{cases} -800 + 1000 + 60 + f_{t+1}(1). & \text{Sell} \\ 80 + f_{t+1}(2). & \text{Keep} \end{cases}$$

$$f_0(0) = 1000 + 60 + f_1(1). \qquad \text{Keep}$$

We should adopt backwards induction:

	S_5	S_4	S_3	S_2	S_1	S_0
$f_t(3)$	-500^{S}	-240^{S}	20^{s}	260^{s}	520^{s}	$-$
$f_t(2)$	-600^{s}	-380^{K}	-120^{K}	$160^{K/S}$	380^{K}	$-$
$f_t(1)$	-800^{S}	-540^{s}	-300^{K}	$-40^{K/S}$	220^{S}	$-$
$f_t(0)$	$-$	$-$	$-$	$-$	$-$	1280^{K}

IPP using DP

Consider $\left\{ \begin{array}{ll} \max & \sum_{i=1}^{n} c_i x_i \\ \text{s.t.} & \sum_{i=1}^{n} a_i x_i \leq W \\ & x_i \geq 0 \ \& \text{ int } \forall i \end{array} \right\}$

Write it as a generalized IPP:

$$\begin{array}{ll} \max & \sum_{t=1}^{T} r_t(x_t) \\ \text{s.t.} & \sum_{t=1}^{T} g_t(x_t) \leq W \\ & x_t \geq 0 \ \& \text{ integer } \forall i \end{array}$$

· We require a DP formulation of the above:

Stages : Represent the product which will be manufactured
{Think of it as if we only produce the t^{th} product x_t in time period t }

States : Represent the amount of resource available

Decision : Number of products x_t produced
{ The t^{th} product costs a_t units of resource and makes C_t units of profit }

$f_{T+1}(d) = 0 \quad \forall \ d$ and for $0 \le t \le T$:

$$f_t(d) = \max_{x_t} \{ r_t(x_t) + f_{t+1}(d - g_t(x_t)) \},$$

$g_t(x_t) \le d$,
x_t integer & $x_t \geq 0$; Report $f_0(w)$

☆ DP has one advantage over IPP ; DP can handle non-linear functions r_i , g_i as well, while IPP can not.

· Up until now, we have known costs / rewards with complete certainty. We now introduce a probabilistic element to DP problems.

Probabilistic DP

If cost / reward of current state is not known with certainty,

OR

Next state in next stage is not known with certainty, then the problem can be solved using <u>Probabilistic DP</u>.

The DP problems before had a recursive relation of the form:

$$f_t(\text{Current State}) = \min/\max \left\{ (\text{Cost/Reward of Current State}) + (f_{t+1}(\text{Next State})) \right\}$$

Can take values w/ some probability

Ex: Current Stage costs uncertain but next period's state is certain

For a price of 1/gallon a supermarket chain has purchased 6 gallons of milk. Each gallon is sold in the chain's 3 stores for 2/gallon. The dairy buys back for 0.5/gallon any left milk at EOD. Demand for each store is uncertain:

The supermarket has to allocate the 6 gallons of milk to the 3 stores so as to maximize the expected net daily profit.

	Daily Demand (gallons)	Probability
Store 1	(1, 2, 3)	(0.6, 0, 0.4)
Store 2	(1, 2, 3)	(0.5, 0.1, 0.4)
Store 3	(1, 2, 3)	(0.4, 0.3, 0.3)

Denote the three stores as the three stages. State at stage t denotes the number of gallons of milk remaining.

$r_t(g_t) :=$ Expected revenue earned from g_t gallons assigned to store t

$f_t(x) :=$ Maximum expected revenue earned from x gallons assigned to stores $t \dots 3$

We note for $t=3$: $f_3(x) = r_3(x)$
and for $t \in \{1, 2\}$:

$$f_t(x) = \max_{0 \le g_t \le x} \{ r_t(g_t) + f_{t+1}(x - g_t) \}$$

$\because$ Demand for each store does not exceed 3, we shall not assign more than 3 gallons to any store. We use backward induction and compute $f_3(\cdot) \longrightarrow f_2(\cdot) \longrightarrow f_1(6)$.
We begin by computing $r_t(g_t)$

$$r_3(3) = \sum_{d=1}^{d=3} p_d \cdot [\, 2 \cdot \overbrace{d}^{\text{Sold}} + 0.5(\overbrace{g_t - d}^{\text{left-over}})] = 4.35 \qquad g_t = 3$$

$$= 0.3 \overbrace{(6+0)}^{d=3} + 0.3 \overbrace{(4+0.5)}^{d=2} + 0.4 \overbrace{(2+1)}^{d=1}.$$

$r_3(2) = 0.3 \times 4 + 0.3 \times 4 + 0.4 \times 2.5 = 3.4$,
$r_3(1) = 0.3 \times 1 + 0.3 \times 1 + 0.4 \times 1 = 1$ &
$r_2(3) = 0.4 \times 6 + 0.1 \times 4.5 + 0.5 \times 3 = 4.35$,
$r_2(2) = 0.4 \times 4 + 0.1 \times 4 + 0.5 \times 3 = 3.5$, $r_2(1) = 2$ &
$r_1(3) = 0.4 \times 6 + 0 + 0.6 \times 3 = 4.2$,
$r_1(2) = 0.4 \times 4 + 0 + 0.6 \times 2.5 = 3.1$, $r_1(1) = 2$

$g_t(x) :=$ Optimal allocation of milk to store t when x units remain in stage t

Stage 3: $f_3(x) = r_3(x)$, $g_3(x) = x$
Need not calculate $f_3(4)$, $f_3(5)$, $f_3(6)$

Stage 2:

$f_2(0) = r_2(0) + f_3(0) = 0$, $g_2(0) = 0$

$$f_2(1) = \max \begin{Bmatrix} r_2(0) + f_3(1) = 2 \\ r_2(1) + f_3(0) = 2 \end{Bmatrix} \Rightarrow f_2(1) = 2, \; g_2(1) \in \{0, 1\}$$

$$f_2(2) = \max \begin{Bmatrix} r_2(0) + f_3(2) = 3.4 \\ r_2(1) + f_3(1) = 4 \\ r_2(2) + f_3(0) = 3.25 \end{Bmatrix} \Rightarrow f_2(2) = 4, \; g_2(1) = 4$$

$$f_2(3) = \max_{0 \le g_t \le 3} \{ r_2(g_t) + f_3(3 - g_t) \} = 5.4, \quad g_2(3) = 1$$

$$f_2(4) = \max_{1 \le g_t \le 3} \{ r_2(g_t) + f_3(4 - g_t) \} = 6.65, \quad g_2(4) = 2$$

Won't allot more than 3 to 2 $\Rightarrow g_t \le 3$

won't allot more than 3 to 3 $\Rightarrow 4 - g_t \le 3 \Rightarrow 1 \le g_t$

$$f_2(5) = \max_{2 \le g_t \le 3} \{ r_2(g_t) + f_3(5 - g_t) \} = 7.75, \quad g_2(5) = 3$$

Same reason as above

$$f_2(6) = r_2(3) + f_3(3) = 8.7, \quad g_2(6) = 3$$

$$f_1(6) = \max_{0 \le g_t \le 3} \{ r_1(g_t) + f_2(6 - g_t) \} = 9.75, \quad g_1(6) \in \{1, 2\}$$

Tracing optimal decisions:

S1 $\rightarrow$ 1 $\rightarrow$ S2 $\rightarrow$ 3 $\rightarrow$ S3 $\rightarrow$ 2

S1 $\rightarrow$ 2 $\rightarrow$ S2 $\rightarrow$ 2 $\rightarrow$ S3 $\rightarrow$ 2

Ex: 3-period inventory problem:

Production cost is $c(x) = 3 + 2x$, if $x > 0$ and 0 otherwise.

Max. production in each period is 4.

After production, the period's random demand is observed. Demand in each period is either 1 or 2 w/ equal probability. Demand must be met on time.

Inventory holding cost is 1/unit.

Max. of 3 units of inventory can be held. Inventory at the end of 3rd period can be sold at 2/unit.

At beginning of 1st period, firm has 1 unit.

Find a production policy to min. expected net cost during the 3 periods.

Stages: $t = 1, 2, 3$ are the 3 time periods

States: Inventory on hand (i)

$f_t(i)$:= Min. expected cost incurred during periods $t, t+1, \dots, 3$

Decision: No. of units to produce.

· We always make sure that the current demand is met, using i & x

Stage 3:

$$f_3(i) = \min_x \left\{ c(x) + \underbrace{\tfrac{1}{2}\cdot(i+x-1)\cdot 1 + \tfrac{1}{2}\cdot(i+x-2)\cdot 1}_{\text{Inventory Cost}} \underbrace{- \tfrac{1}{2}(i+x-1)\cdot 2 - \tfrac{1}{2}\cdot(i+x-2)\cdot 2}_{\text{Salvage Value}} \right\}$$

$x \leq 4 \Leftarrow$ Production limit

$i + x \geq 2 \Leftarrow$ Demand must be met

$i + x - 1 \leq 3 \Leftarrow$ Inventory has a limit

Similarly, for $i \in \{1, 3\}$

i	x	$c(x)$	Inventory Cost $i+x-\frac{3}{2}$	Salvage Value $2(i+x)-3$	Total Expected Cost	$f_3(i), x_3(i)$
0	2	7	1/2	1	6.5	$f_3(0)=6.5$
$2 \leq x \leq 4$	3	9	3/2	3	7.5	$x_3(0)=2$
	4	11	5/2	5	8.5	
2	0	0	1/2	1	-0.5	$f_3(2)=-1/2$
$0 \leq x \leq 2$	1	5	3/2	3	3.5	$x_3(2)=0$
	2	7	5/2	5	4.5	

Stage 2:

$$f_2(i) = \min_x \left\{ c(x) + \tfrac{1}{2}(i+x-1)\cdot 1 + \tfrac{1}{2}(i+x-2)\cdot 1 + \tfrac{1}{2}\cdot f_3(i+x-1) + \tfrac{1}{2}\cdot f_3(i+x-2) \right\}$$

$i + x \geq 2, \quad i+x-1 \leq 3, \quad x \leq 4$

Table will have the same columns, except that "Salvage Cost" is replaced by "$[f_3(i+x-1) + f_3(i+x-2)]/2$"

Stage 1: More or less the same as S2:

$$f_1(i) = \min_x \left\{ c(x) + \tfrac{1}{2}(i+x-1)\cdot 1 + \tfrac{1}{2}(i+x-2)\cdot 1 + \tfrac{1}{2}\cdot f_2(i+x-1) + \tfrac{1}{2}\cdot f_2(i+x-2) \right\}$$

$i + x \geq 2, \quad i+x-1 \leq 3, \quad x \leq 4, \quad i = 1$

Now that we have the table filled with optimal values of $f_t(i)$ and optimal decisions $x_t(i)$, we can trace the optimal policy for the firm.

Markov Decision Process

Essentially PDP but with a long infinite time horizon.

T denotes the number of time periods or planning horizon.

In MDPs, T is very large.

Time value of money:

1 unit of money one period from now $= \beta \times 1$ unit of money now $(0<\beta<1)$

Similarly,

1 unit of money two periods from now $= \beta^2 \times 1$ units of money now, $0<\beta<1$

Some terminologies related to MDP:

1) State Space: At the beginning of each period, the MDP is in some state. Denote $S = \{1, 2, \ldots, N\}$ as the set of all possible states. S is called the state space.

2) Decision Set: For each state i, there is a finite set of allowable decisions $D(i)$

3) Transition Probabilities: Suppose a period begins in state i and $d \in D(i)$ is chosen. $\mathbb{P}[j \mid i, d]$ is the probability that the next period's stage is j.

4) Expected Reward: During a period in which state is i & decision is $d \in D(i)$ and expected reward $r_{i,d}$ is received.

Usually in the current period

Eg: At the beginning of each week, a machine is in one of 4 conditions (states): E G A B (Excellent, Good, Average, Bad).

Weekly revenue earned by a machine in each type of condition is as follows: (100, 80, 50, 10). After observing the condition at the beginning of the week, we have the option to instantaneously replace with excellent machine which costs 200. Quality of machine deteriorates over time as:

Present State	E	G	A	B
E	0.7	0.3		
G		0.7	0.3	
A			0.6	0.4
B				1

Prob. of state in next week

We need to determine the state space, decision sets, transition probabilities & expected rewards.

State Space: $\{E, G, A, B\}$ — the state of the machine.

Decision Set: R — to Replace
NR — Not to Replace

Decision in each state:
$D(E) = \{NR\}, \quad D(G) = D(A) = D(B) = \{R, NR\}$

Transition Probabilities: $\mathbb{P}[j / i, d]$

$\mathbb{P}[E \mid E, NR] = 0.7, \quad \mathbb{P}[E \mid G, NR] = 0,$
$\mathbb{P}[G \mid E, NR] = 0.3, \quad \mathbb{P}[G \mid G, NR] = 0.7,$
$\mathbb{P}[A \mid E, NR] = 0, \quad \mathbb{P}[A \mid G, NR] = 0.3,$
$\mathbb{P}[B \mid E, NR] = 0, \quad \mathbb{P}[B \mid G, NR] = 0,$
$\mathbb{P}[E \mid A, NR] = 0, \quad \mathbb{P}[E \mid B, NR] = 0,$
$\mathbb{P}[G \mid A, NR] = 0, \quad \mathbb{P}[G \mid B, NR] = 0,$
$\mathbb{P}[A \mid A, NR] = 0.6, \quad \mathbb{P}[A \mid B, NR] = 0,$
$\mathbb{P}[B \mid A, NR] = 0.4, \quad \mathbb{P}[B \mid B, NR] = 1,$

$\because$ Replacement takes place instantaneously & we are talking about the next week

$\mathbb{P}[E \mid G, R] = 0.7 = \mathbb{P}[E \mid A, R] = \mathbb{P}[E \mid B, R],$
$\mathbb{P}[G \mid G, R] = 0.3 = \mathbb{P}[G \mid A, R] = \mathbb{P}[G \mid B, R],$
$\mathbb{P}[A \mid G, R] = 0 = \mathbb{P}[A \mid A, R] = \mathbb{P}[A \mid B, R],$
$\mathbb{P}[B \mid G, R] = 0 = \mathbb{P}[B \mid A, R] = \mathbb{P}[B \mid B, R],$

Expected Rewards: $r_{i,d}$

$r_{E,NR} = 100, \quad r_{G,NR} = 80, \quad r_{A,NR} = 50, \quad r_{B,NR} = 10$
$r_{G,R} = -200 + 100 = -100 = r_{B,R} = r_{A,R}$

☆ A **Policy** is a rule which describes how a period's decision is chosen.

☆ A policy δ is a **Stationary Policy** if whenever the state is i, the policy chooses independently of the period, the same decision. We call this decision δ.

δ: arbitrary policy

Δ: Set of all policies

We shall only concern ourselves with stationary policies in this course.

X_1: Given state of the MDP at the beginning of period 1 (initial state)

X_t: Random Variable for the states of the MDP beginning at period t, $t: 2, 3, \ldots$

d_t: Decision chosen in time t

$v_\delta(i)$: Expected discounted reward earned during an infinite number of periods given that at the beginning of period 1, state is i & stationary policy is δ

- In a maximization problem,

$$v(i) := \max_{\delta \in \Delta} v_\delta(i)$$

- If a policy δ^* is such that

$$v(i) = v_{\delta^*}(i) \quad \forall \quad i \in S$$

then, δ^* is an optimal policy.

Discount Factor

- Now, to find $v_\delta(i)$ w/ $0<\beta<1$:

$$v_\delta(i) = r_{i,\delta(i)} + \beta \cdot \sum_{j\in S} \mathbb{P}[j \mid i, \delta(i)] \cdot v_\delta(j) \quad \forall\ i \in S$$

These equations of $v_\delta(i)$ are called Value Determination Equations.

We now discuss some methods to find δ^*:

Described below for maximization problems

Howard's Policy Iterative Method

Step 1: Find $v_\delta(i)$ $\forall$ $i \in S$

S2: $T_\delta(i) := \max\limits_{d\in D(i)} \{ r_{i,d} + \beta \sum\limits_{j\in S} \mathbb{P}[j \mid i, d]\, v_\delta(j) \}$

If $T_\delta(i) = v_\delta(i)$ $\forall$ $i \in S$ then δ is an optimal policy. Terminate.

If $T_\delta(i) > v_\delta(i)$ for some i then δ is not optimal.

S3: Construct another policy δ' which is modified from δ so that the decision in i is the decision which attains $T_\delta(i)$.

Return to S1 by replacing δ w/ δ'

Ex: Back to the previous example to find the optimal policy w/ $\beta = 0.9$:

Consider the following policy:

$\delta(E) = \delta(G) = NR, \quad \delta(A) = \delta(B) = R$

Onto Step 1:

$v_\delta(E) = 100 + 0.9\,[0.7\, v_\delta(E) + 0.3\, v_\delta(G)]$
$v_\delta(G) = 80 + 0.9\,[0.7\, v_\delta(G) + 0.3\, v_\delta(A)]$
$v_\delta(A) = -100 + 0.9\,[0.7\, v_\delta(E) + 0.3\, v_\delta(G)]$
$v_\delta(B) = -100 + 0.9\,[0.7\, v_\delta(E) + 0.3\, v_\delta(G)]$

Solve to get

$v_\delta(E) = 687.81$, $v_\delta(G) = 572.19$,
$v_\delta(A) = 487.81 = v_\delta(B)$

Onto S2: $T_\delta(E) = v_\delta(E) = 687.81$,

$T_\delta(G) = \max\{ v_\delta(G) = (NR \text{ at } G) = 572.19,\ (R \text{ at } G) = -100 + 0.9[0.7\, v_\delta(E) + 0.3\, v_\delta(G)] = 487.81 \} = v_\delta(G)$

$T_\delta(A) = \max\{ (NR \text{ at } A) = 50 + 0.9[0.6\, v_\delta(A) + 0.4\, v_\delta(B)] = 489.03,\ (R \text{ at } A) = v_\delta(A) = 487.81 \} = 489.03 > v_\delta(A)$

$T_\delta(B) = v_\delta(B)$

Onto S3:

$\delta'(A) = NR$ & $\delta'(i) = \delta(i)\ \forall\, i \in \{E, G, B\}$

Calculate $v_{\delta'}(i)\ \forall\, i \in S$, $T_{\delta'}(i)$ and check for optimality of δ'. It will turn out to be, in fact, optimal.

Solving the Value Determination Equation for δ', we obtain:

$v_{\delta'}(E) = 690.23$, $v_{\delta'}(G) = 575.5$,
$v_{\delta'}(A) = 492.35$, $v_{\delta'}(B) = 490.23$

& $T_{\delta'}(i) = v_{\delta'}(i)\ \forall\, i \in S$ for

$(E, G, A, B) \xrightarrow{\delta'} (NR, NR, NR, R)$

LP Approach to find Optimal Policy

for a maximization problem

$$v_\delta(i) = r_{i,d} + \beta \cdot \sum_{j \in S} \mathbb{P}[j \mid i, d] \cdot v_\delta(j), \quad d \in D(i)$$

We require such a δ for which the LHS is always greater than or equal to RHS for all decisions $d \in D(i)$, i.e.,

$$v_\delta(i) - \beta \sum_{j \in S} \mathbb{P}[j \mid i, d] \cdot v_\delta(j) \geq r_{i,d} \quad \forall\, d \in D(i)\ \forall i$$
Constraints

$$\min_{\delta \in \Delta} \sum_{i \in S} v_\delta(i)$$
Objective Function

min used for a maximization problem

$v_\delta(i)$: u.r.s. decision variables

☆ Finding optimal $v_\delta(i)$ from optimal solution of LP is trivial.
To construct δ, look at the d's from the binding constraints, i.e., the constraints for which equality holds. Ties can be broken arbitrarily.

Eg: For the above example, the LP formulation is as follows:

		Equality Holds	Corresponding Optimal decisions
min.	$V_E + V_G + V_A + V_B$		
s.t.	$V_E \geq 100 + 0.9(0.7\, V_E + 0.3\, V_G)$	$\longleftrightarrow$	NR
	$V_G \geq 80 + 0.9(0.7\, V_G + 0.3\, V_A)$	$\longleftrightarrow$	NR
	$V_G \geq -100 + 0.9(0.7\, V_E + 0.3\, V_G)$		R
	$V_A \geq 50 + 0.9(0.6\, V_A + 0.4\, V_B)$	$\longleftrightarrow$	NR
	$V_A \geq -100 + 0.9(0.7\, V_E + 0.3\, V_G)$		R
	$V_B \geq 10 + 0.9(1\, V_B)$		NR
	$V_B \geq -100 + 0.9(0.7\, V_E + 0.3\, V_G)$	$\longleftrightarrow$	R

V_E, V_G, V_A, V_B u.r.s. decision variables

☆ For a minimization DP problem, the objective function of the LP formulation is to be maximized.
sum of decision variables

Markov Chains

- Deals with Stochastic Processes (How Random Variables change over time)
- Suppose we observe some characteristics of the system at discrete periods of time, say, $t = 0, 1, \ldots$

  Let X_t be the value of the characteristic at time t. Many times X_t is an R.V.

  A discrete time stochastic process is a description of the relation b/w the RV's $X_0, X_1, \ldots$
- Time can be discrete or continuous. Likewise, states (X_t) can be discrete or continuous. Thus, there are, broadly speaking, 4 kinds of Stochastic Processes.

Gambler's Ruin

At time 0, you have ₹2. At times $1, 2, \ldots,$ suppose you play a game in which you bet Re 1 and with prob. p you win the game & with prob. $1-p$ you lose the game.

Your goal is to increase your capital to ₹4 & as soon as you do so, the game is over. The game is also over if your capital is reduced to 0.

$X_t :=$ Your capital position after time t. Game starts at $t = 1$.

Then $X_0, X_1, X_2, \ldots$ constitute a discrete time Stochastic Process.

Balls from an Urn

An urn contains 2 unpainted balls at present. We choose a ball at random & flip a coin. If the chosen ball is unpainted & coin shows H, we paint the chosen ball red. If the chosen ball is unpainted & the coin shows T, we paint the chosen ball black. If the chosen ball is painted, then regardless of what the coin shows, we change the colour of the ball.

To model this situation as a stochastic process, we define time t to be the time after the coin has been tossed t times & the chosen ball has been painted.

State at any time may be described by the vector $[u, r, b]$, where u is the number of unpainted balls in the urn, r is the no. of red balls in the urn, & b is the no. of black balls in the urn.

Note $X_0 = [2, 0, 0]$, $X_1 = [1, 0, 1]$ or $X_1 = [1, 1, 0]$.

☆ A continuous time stochastic process is a stochastic process in which the state of the system can be viewed at any time.

Ex: The number of people in a supermarket t minutes after the store opens is a cts. time stochastic process.

Def: A discrete time stochastic process is a **Markov Chain** if $\forall\ t \geq 0\ \ \forall\ i \in S$, we have:

$$\mathbb{P}[X_{t+1} = i_{t+1} \mid X_0 = i_0, X_1 = i_1, \dots, X_t = i_t] = \mathbb{P}[X_{t+1} = i_{t+1} \mid X_t = i_t]$$

☆ A Markov Chain is a stochastic process with a distinct "lack of memory".

Def: In a Markov Chain, suppose that $\forall\ i, j\ \ \forall\ t$:

$$\mathbb{P}[X_{t+1} = j \mid X_t = i] = \mathbb{P}[i, j] = p_{ij}$$

i.e., $\mathbb{P}[X_{t+1} = j \mid X_t = i]$ is independent of t, where p_{ij} is the probability that given the system is in state i at time t, it will be in state j at time $t+1$ (for all time t).

Such a Markov Chain is known as a **Stationary Markov Chain**.

Def: If a system moves from state i during one period to state j during the next period, we say that a **transition** from i to j has occurred.

Note: p_{ij}'s are called the **transition probabilities** of a Markov Chain.

Note: Suppose $S = \{1, 2, \dots, s\}$ is the set of possible states. The transition probabilities are represented by an $s \times s$ matrix P:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}_{s \times s}$$

Observe: Row-wise sum of each row is equal to 1

This is called Transition Probability Matrix.

Note: Given that at time t, the state is i, the process must be in some state so that we have,

$$\sum_{j \in S} \mathbb{P}[X_{t+1} = j \mid X_t = i] = 1 \Rightarrow \sum_{j \in S} p_{ij} = 1 \quad \forall\ i \in S$$

Ex: Gambler's Ruin problem with winning prob p

$S = \{0, 1, 2, 3, 4\}$

$$P = \begin{array}{c} \overbrace{\begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array}}^{S} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-w & 0 & w & 0 & 0 \\ 0 & 1-w & 0 & w & 0 \\ 0 & 0 & 1-w & 0 & w \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \left.\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\right\} S \qquad 5 \times 5$$

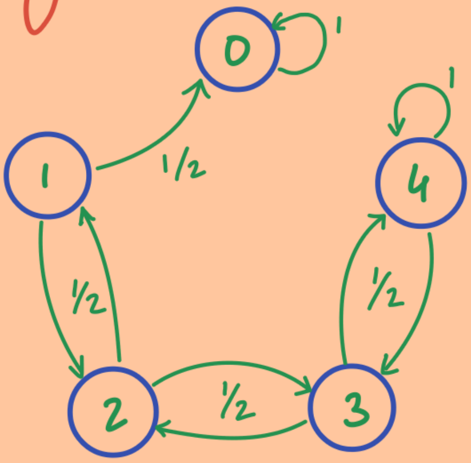
Ex: Balls from an urn with 2 unpainted balls initially.

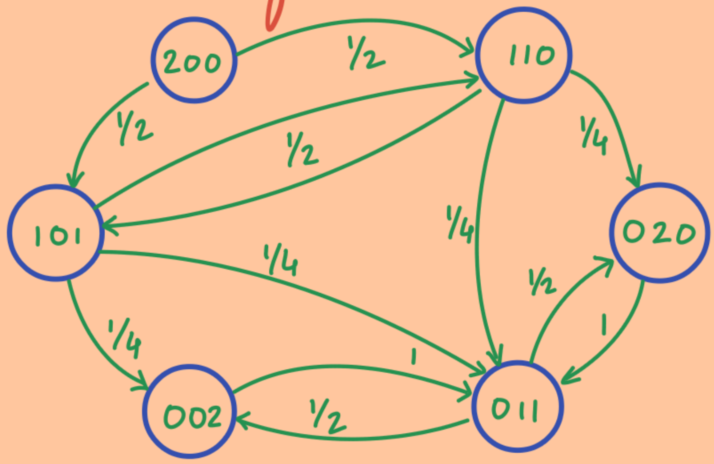
$S = \{[2,0,0], [1,0,1], [1,1,0], [0,2,0], [0,0,2], [0,1,1]\}$

$$\begin{array}{c|cccccc} & \multicolumn{6}{c}{S} \\ & 200 & 110 & 101 & 020 & 002 & 011 \\ \hline [2\,0\,0] & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ [1\,1\,0] & 0 & 0 & 1/2 & 1/4 & 0 & 1/4 \\ [1\,0\,1] & 0 & 1/2 & 0 & 0 & 1/4 & 1/4 \\ [0\,2\,0] & 0 & 0 & 0 & 0 & 0 & 1 \\ [0\,0\,2] & 0 & 0 & 0 & 0 & 0 & 1 \\ [0\,1\,1] & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{array} \qquad 6 \times 6$$

☆ Transition probabilities can also be expressed in the form of a graph (generally directed). Can also be thought of as a Finite State Automata:

Gambler's Ruin, Balls from an Urn





n-step Transition Probabilities

If a Markov Chain is in state i at time m, we are interested in the probability that n periods later, the Markov chain will be in state j.

Call this probability $p_{ij}(n)$ since this probability will be independent of m for Stationary Markov Chains. $p_{ij}(n)$ is called n-step transition probability from state i to state j.

Obviously, $p_{ij}(1) = p_{ij}$. For $p_{ij}(2)$, we enumerate all possibilities $i \longrightarrow k \longrightarrow j$:

$$p_{ij}(2) = \sum_{k \in S} p_{ik} \cdot p_{kj} = (P^2)_{ij}$$

And in general, we have $\boxed{p_{ij}(n) = (P^n)_{ij}}$

Also, there is the Chapman-Kolmogorov Identity:

$$\boxed{p_{ij}(l+m) = \sum_{k \in S} p_{ik}(l) \cdot p_{kj}(m)}$$

Ex: Suppose the entire Cola industry produces only 2 Colas. Given that a person last purchased Cola 1, there is a 90% chance that her next purchase will be Cola 1. Given that a person last purchased Cola 2, there is an 80% chance that her next purchase will be Cola 2.

Each person's purchase is a Markov Chain. State at any time t is the Type of cola the person just purchased.

State 1: Person just purchased Cola 1

State 2: Person just purchased Cola 2

And the transition matrix is

$$P = \begin{array}{r} \\ \text{Cola 2} \\ \\ \end{array} \begin{array}{c} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \\ \begin{array}{cc} \text{Cola 1} & \text{Cola 2} \end{array} \end{array}$$

Let X_0 be the present purchase.
& X_n be the purchase in the n^{th} future period from now. Then, $X_0, X_1, \dots, X_n$ represents a Markov Chain

Q: If a person is currently a Cola 2 purchaser, what is the probability that she will purchase Cola 1 two purchases from now?

Required to find $(2,1)^{th}$ element of P^2

$$P^2 = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

Required probability is $\underline{0.34}$ $\leftarrow P_{21}^{(2)}$

Q: If a person is currently a Cola 1 purchaser, what is the probability that she will purchase Cola 2 three purchases from now?

Required to find $(1, 2)^{th}$ element of P^3

$$P^3 = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix}$$

Required probability is $\underline{0.219} = p_{12}^{(3)}$

Note: We define q_i to be the probability that the chain is in state i at time $t=0$, i.e. $\mathbb{P}[X_0 = i] = q_i$.

Define $q = [q_1 \ q_2 \dots q_s]$ as the initial probability distribution for the Markov Chain.

☆ In such a case, the probability that the system is in state j at time t is given by $\sum_{k \in S} q_k \cdot P_{kj}(t)$

OR

in other words, $[q_1 \ q_2 \dots q_s] \cdot [p_{1j}(n) \dots p_{sj}(n)]$

Eg: In the Cola example, suppose that 60% of the people drink Cola 1 & 40% of the people drink Cola 2. Three purchases from now, what fraction of people will be drinking Cola 1?

$q = [0.6 \ \ 0.4]$; Need to find:

$$\sum_{k \in S} q_k P_{k1}(3) = q_1 P_{11}(3) + q_2 P_{21}(3)$$

$$= (q_1 \ \ q_2) \begin{pmatrix} P_{11}(3) \\ P_{21}(3) \end{pmatrix} = (0.6 \ \ 0.4) \begin{pmatrix} 0.781 \\ 0.438 \end{pmatrix} = 0.6438$$

Classification of States in a Markov Chain

Def: State j is said to be accessible from state i if $P_{ij}(n) > 0$ for some $n \geq 0$ Equality does not make sense for $i \neq j$

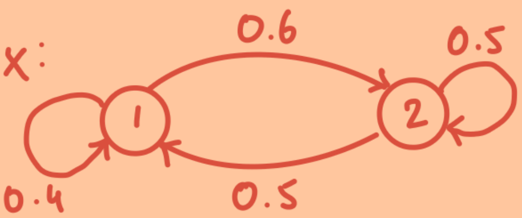
Def: Two states i & j that are accessible from each other are said to communicate

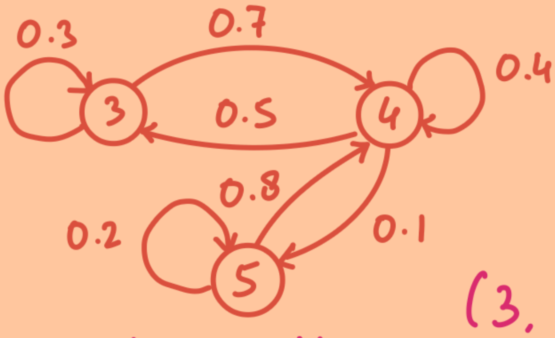
Note: The relation of communication is an equivalence relation.
This relation will be used to induce a partition on the set of states
into different classes

Def: Two states that communicate are said to be in the same class.

Def: A Markov Chain is said to be irreducible if there is only one class, i.e., all states communicate with each other.

Ex:





(1, 2) are in the same class.
(3, 4, 5) are also in the same class, but different from that of (1, 2).
Markov Chain is not irreducible $\because$ all states do not communicate w/ one another.

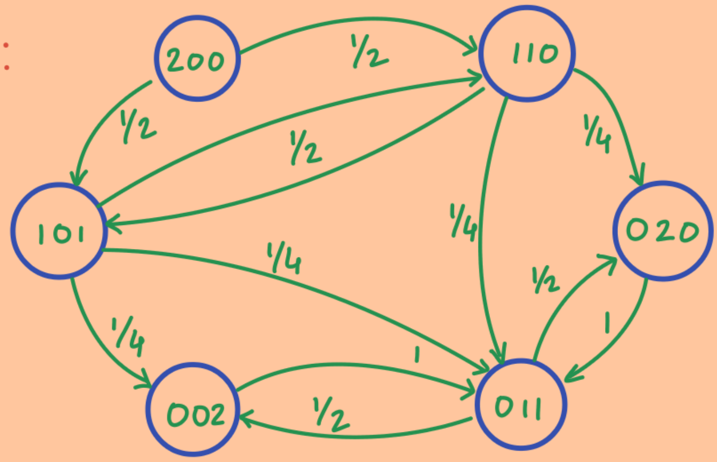
Def: A state i is an absorbing state if $P_{ii} = 1$, i.e., whenever we enter an absorbing state we never leave the state.

Def: For any state i, let f_i denote the probability that starting in state i, the process will ever re-enter state i.

State i is said to be recurrent if $f_i = 1$ & transient if $f_i < 1$

To check if a state i is recurrent, verify if $\exists$ a path to come back to i (eventually) for each possible transition or not. If $\exists$ even a single state j s.t. j is accessible from i but i is not accessible back from j, then the state i is transient, not recurrent.

Eg:



200, 110, 101 are transient states.
011, 020, 002 are recurrent states.

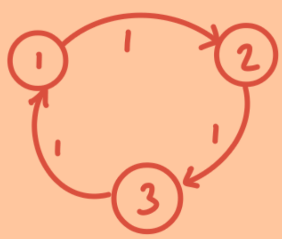
Theorem: State i is recurrent if $\sum_{n=1}^{\infty} P_{ii}(n) = \infty$ and transient if $\sum_{n=1}^{\infty} P_{ii}(n) < \infty$.

Note: If state i is recurrent and communicates with state j, then state j is also recurrent.

Def: A state i is periodic with period $d > 1$ if d is the smallest number such that only in multiples of d steps, the system returns to state i.

If a recurrent state is not periodic, it is referred to as aperiodic.

Eg: Consider the Markov Chain



$\rightarrow$ All states are periodic w/ period 3

Def: If all states in a chain are recurrent, aperiodic and communicate with each other, (irreducible), then the chain is called ergodic.

Ex: $$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1/3 & 2/3 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/4 & 3/4 \end{bmatrix} \end{array}$$

$\rightarrow$ 1 & 3 are aperiodic because they can return to themselves in $d=1$ steps

2 is not periodic because chain can return in 2, 3, ... steps. So, there is no d having all the numbers as multiples.

Eg: $$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1/4 & 3/4 \end{bmatrix}$$ $\longrightarrow$ Chain is not ergodic.

Steady State Probabilities

Theorem: If P is a transition matrix for an s-state ergodic chain, $\exists$ a vector $\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_s]$ s.t.

$$\lim_{n \to \infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_s \\ \pi_1 & \pi_2 & \ddots & \vdots \\ \vdots & \vdots & & \\ \pi_1 & \pi_2 & \dots & \pi_s \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix}$$

- The above theorem implies that the Markov chain settles down independently of the initial state i & there is a probability π_j that we are in state j.

Def: $\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_s]$ is called the steady-state distribution or equilibrium distribution of a Markov Chain.

Corollary: For large n, we have,

$$P_{ij}(n+1) \approx P_{ij}(n) \approx \pi_j$$

Also, we have,

$$P_{ij}(n+1) = \sum_{k\in S} P_{ik}(n) \cdot P_{kj}(1) = \sum_{k \in S} \pi_k \cdot P_{kj}$$

$\therefore$ We get,

$$\pi_j = \sum_{k\in S} \pi_k \cdot P_{kj}$$

$$\Longleftrightarrow \boxed{\pi = \pi P}$$

- There are s equations with s unknowns.
- Need rank of P to be s for a unique solution. However, rank of P can be at most $(s-1)$ {$\because$ row-wise sum is 1, $\therefore$ if we are given $(s-1)$ columns, we can find the last column. Hence, they're not linearly independent.}.

  Thus, one equation is redundant. Also, there exist infinite solutions to $\pi = \pi P$
- Add the constraint $\pi_1 + \pi_2 + \ldots + \pi_s = 1$ and solve to get a unique solution.

  $p_{i1}(n) + p_{i2}(n) + \ldots + p_{is}(n) = 1 \longrightarrow \lim(n \to \infty)$
- Thus to obtain π, we can solve for the following equations: $\sum_{i\in S} \pi_i = 1, \quad \pi = \pi P$

Ex: Steady State Probability distribution of Cola example $\quad P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$

$$\pi = \pi P \Rightarrow [\pi_1 \;\; \pi_2] = [\pi_1 \;\; \pi_2] \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = [0.9\pi_1 + 0.2\pi_2, \; 0.1\pi_1 + 0.8\pi_2]$$

$$\Rightarrow \pi_1 = 0.9\pi_1 + 0.2\pi_2 \quad \Rightarrow 0.1\pi_1 = 0.2\pi_2, \quad 0.2\pi_2 = 0.1\pi_1$$
$$\pi_2 = 0.1\pi_1 + 0.8\pi_2$$

$\Rightarrow \pi_1 = 2\pi_2, \quad \pi_1 + \pi_2 = 1 \Rightarrow \underline{\pi_2 = 1/3}, \quad \underline{\pi_1 = 2/3}$

☆ $\pi = \pi P \Rightarrow \pi_j = \sum_{k \in S} \pi_k P_{kj} \longrightarrow$ Subtract $\pi_j \cdot p_{jj}$ from both sides

$$\underbrace{\pi_j (1 - p_{jj})}_{(1)} = \underbrace{\sum_{j \neq k \in S} \pi_k \cdot P_{kj}}_{(11)} \Rightarrow (1) = (2)$$

(1) Probability that we leave j, given that we are at state j right now.

(11) Probability that we enter j, given that we are not at state j right now.

Ex: Cola example: Suppose that each customer makes one purchase per week (52 weeks in a year). Suppose there are 100M customers.

One selling unit of Cola costs the company 1 unit to produce & sells for 2 units.

For 500 million per year, an advertising firm guarantees to decrease from 10% to 5% the fraction of Cola 1 users who switch to Cola 2 after a purchase.

Should the Cola 1 company hire the advertisement firm?

Previously,

$$\pi_1 = 2/3, \quad \pi_2 = 1/3$$

Profit per year: $2/3 \cdot 52 \cdot 100{,}000{,}000 \cdot 1$
$= 3{,}466{,}666{,}667$

Now,

$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \longrightarrow \pi = \pi P \rightsquigarrow \pi_1 = 4\pi_2,$$

$$\pi_1 + \pi_2 = 1 \Rightarrow \pi_1 = 4/5$$

Profit per year: $4/5 \cdot 52 \cdot 100{,}000{,}000 \cdot 1 - 500{,}000{,}000 =$

$3{,}466{,}666{,}667 < 3{,}660{,}000{,}000$

Thus, the Cola 1 company should hire the advertising firm

Mean First Passage Time

For an Ergodic Markov Chain, let

$m_{ij} :=$ Expected number of transitions before we first reach state j, given that we are currently in state i

m_{ij} is called the mean first passage time from state i to state j.

$$\boxed{m_{ij} = 1 \cdot p_{ij} + \sum_{j \neq k \in S} p_{ik} \cdot (1 + m_{kj})}$$

$$= 1 + \sum_{j \neq k \in S} p_{ik} \cdot m_{kj}$$

Ex: Find m_{12}, m_{21} in Cola Example:

$m_{12} = 1 + p_{11} m_{12} \Rightarrow m_{12} = 1/(1-p_{11}) = 1/(1-0.9) = 10$

$m_{21} = 1 + p_{22} m_{12} \Rightarrow m_{21} = 1/(1-p_{22}) = 1/(1-0.8) = 5$

$m_{11} = 1 + p_{12} \cdot m_{21} \Rightarrow m_{11} = 1 + 0.1 \times 5 = 1.5 = \frac{1}{\pi_1} = \frac{1}{2/3}$

Similarly,

$m_{22} = 1 + p_{21} \cdot m_{12} = 1 + 0.2 \times 10 = 3 = 1/\pi_2 = 1/(1/3)$

* Even for more generalized Ergodic Markov Chains, $m_{ii} = \frac{1}{\pi_i}$ holds. This is because eventually, the probability of reaching i is π_i, i.e., success probability of a Geometric R.V. is $\pi_i \Rightarrow$ Expectation $= 1/\pi_i$

Absorbing Chains

Markov Chains in which some of the states are absorbing and the remaining are transient states are called absorbing chains.

Eg: (1) The accounts receivable situation of a firm is often modelled as an absorbing MC.

Suppose a firm assumes that an account is uncollectable if it is more than 3 months overdue. Then at the beginning of each month, each account may be classified into one of the following states:

S1: New account
S2: Payment on account is one month overdue
S3: " " " " two months "
S4: " " " " three " "
S5: Account has been paid
S6: Account is written off as bad debt

The transition matrix from past date is given as:

$$P = \begin{array}{c} \\ \left[\begin{array}{cccc|cc} 0 & 0.6 & 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.3 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \\ \begin{array}{cccc|cc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \end{array} \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}$$

Transient ↑ ; ↓ Absorbing

S1, 2, 3, 4 are transient while S5, 6 are absorbing.

Thus, the MC is absorbing.

(2) A law firm employs 3 types of lawyers: junior lawyers, Senior lawyers & Partners. During a given year, there is a 0.15 probability that a junior lawyer will

be promoted to senior lawyer & 0.05 probability that he/she will leave the firm. Also, there is a 0.2 probability that a senior lawyer will be promoted to a partner and 0.1 probability that he/she will leave the firm. A partner leaves w/ probability 0.05. Model the situation as a MC.

States: Junior, Senior, Partner, Leaving as a Partner, Leaving as a Junior/Senior

$$P = \begin{array}{c} \\ \left[\begin{array}{ccc|cc} 0.8 & 0.15 & 0 & 0 & 0.05 \\ 0 & 0.7 & 0.2 & 0 & 0.1 \\ 0 & 0 & 0.95 & 0.05 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} J \\ S \\ P \\ L_P \\ L_{NP} \end{array} \\ \begin{array}{ccccc} J & S & P & L_P & L_{NP} = L_{J,S} \end{array} \end{array}$$

J, S, P are transient while L_P, L_{NP} are absorbing.

☆ Suppose there are s states in an absorbing chain. Let there be m absorbing states and $s-m$ transient states. Then the transition matrix can be written more generally as

$$\begin{array}{c} \begin{array}{cc} \overbrace{\qquad\quad}^{s-m} & \overbrace{\quad}^{m} \end{array} \\ \begin{array}{c} s-m\{ \\ m\{ \end{array} \left[\begin{array}{c|c} Q & R \\ \hline O & I \end{array}\right] \end{array}$$

Note: If we are at present in transient state t_i, the expected number of periods that will be spent in transient state t_j before absorption is the $(ij)^{th}$ element of the matrix $\underline{(I-Q)^{-1}}$ ⟶ MC's Fundamental Matrix

Kronecker-Delta, we are spending one period if $i=j$

$i, j, k \in [s-m]$ | $(I-Q)^{-1} = I + Q + Q^2 + \ldots$

$$s_{ij} = \delta_{ij} + \sum_{k\in[s-m]} p_{ik}s_{kj} \Rightarrow s_{ij} = \delta_{ij} + \sum_{k\in[s-m]} q_{ik}s_{kj} \Rightarrow S = I + QS \Rightarrow S = (I-Q)^{-1}$$

Note: If we are at present in transient state t_i, the probability that we will eventually be absorbed into absorbing state a_j is the $(ij)^{th}$ element of the matrix $(I-Q)^{-1}R$

$(I-Q)^{-1}R = R + QR + Q^2R + Q^3R + \dots$

$i, k \in [s-m],\ j \in [m]$ So, $p_{ik} = q_{ik}$, $p_{ij} = r_{ij}$

$$\underbrace{s_{ij} = p_{ij} + \sum_{k \in [s-m]} p_{ik} s_{kj}} \Rightarrow s_{ij} = r_{ij} + \sum_{k \in [s-m]} q_{ik} s_{kj} \Rightarrow S = R + QS \Rightarrow S = (I-Q)^{-1}R$$

Q: 1) What is the probability that a new account will be collected eventually.

2) What is the probability that a one-month overdue account will eventually become a bad debt?

3) If the firm's sales average 1,00,000 pm. how much money per year will go uncollected?

$$Q = \begin{pmatrix} 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0.4 & 0 \\ 0.5 & 0 \\ 0.6 & 0 \\ 0.7 & 0.3 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I - Q = \begin{pmatrix} 1 & -0.6 & 0 & 0 \\ 0 & 1 & -0.5 & 0 \\ 0 & 0 & 1 & -0.4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow (I-Q)^{-1} = \begin{array}{c c} \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ t_4 \end{array} & \begin{pmatrix} 1 & 0.6 & 0.3 & 0.12 \\ 0 & 1 & 0.5 & 0.2 \\ 0 & 0 & 1 & 0.4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ & \begin{array}{cccc} t_1 & t_2 & t_3 & t_4 \end{array} \end{array}$$

$$(I-Q)^{-1}R = \begin{array}{c c} & \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ t_4 \end{array} & \begin{pmatrix} 0.964 & 0.036 \\ 0.94 & 0.06 \\ 0.88 & 0.12 \\ 0.7 & 0.3 \end{pmatrix} \end{array}$$

$(1) = (t_1\, a_1)^{th} = (11)^{th} = 0.964$

$(2) = (t_2\, a_2)^{th} = (22)^{th} = 0.06$

Entries of $(I-Q)^{-1}R$

(3) Prob. of new account being absorbed in a_2 is $= 0.036$

& Annual sales are 12,00,000

$\therefore$ Uncollected money $= 12,00,000 * 0.036$
$= 43,200$

Q: 1) What is the average length of time a newly hired junior lawyer spends working for the firm?

2) What is the probability that a junior lawyer makes it to partner?

3) What is the average length of time that a partner spends with a firm (as a partner)?

$$P = \begin{array}{c|ccc|cc} & J & S & P & L_P & L_{J,S} \\ \hline J & 0.8 & 0.15 & 0 & 0 & 0.05 \\ S & 0 & 0.7 & 0.2 & 0 & 0.1 \\ P & 0 & 0 & 0.95 & 0.05 & 0 \\ \hline L_P & 0 & 0 & 0 & 1 & 0 \\ L_{J,S} & 0 & 0 & 0 & 0 & 1 \end{array} \longrightarrow Q = \begin{pmatrix} 0.8 & 0.15 & 0 \\ 0 & 0.7 & 0.2 \\ 0 & 0 & 0.95 \end{pmatrix}$$

$$R = \begin{pmatrix} 0 & 0.05 \\ 0 & 0.1 \\ 0.05 & 0 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I - Q = \begin{pmatrix} 0.2 & -0.15 & 0 \\ 0 & 0.3 & -0.2 \\ 0 & 0 & 0.05 \end{pmatrix} \rightsquigarrow (I-Q)^{-1} = \begin{array}{c} J \\ S \\ P \\ \\ \end{array}\begin{array}{c} \begin{pmatrix} 5 & 2.5 & 10 \\ 0 & 10/3 & 40/3 \\ 0 & 0 & 20 \end{pmatrix} \\ \begin{matrix} J & S & P \end{matrix} \end{array}$$

$$(I-Q)^{-1} R = \begin{array}{c} J \\ S \\ P \\ \\ \end{array}\begin{array}{c} \begin{pmatrix} 0.5 & 0.5 \\ 1/3 & 2/3 \\ 0 & 1 \end{pmatrix} \\ \begin{matrix} L_P & L_{NP} \end{matrix} \end{array}$$

1) Avg. time spent as J +
 " " " " S +
 " " " " P
$= 5 + 2.5 + 10 = 17.5$

2) $(11)^{th}$ entry of $(I-Q)^{-1}R = (J\ L_P)^{th} = 0.5$

3) $(33)^{th}$ element of $(I-Q)^{-1} = (S\ S)^{th} = 20$

Non-Linear Programming

General Non-Linear Programming -

To find decision variables $x_1, x_2, \ldots, x_n$ which

$$\begin{aligned} \max/\min \quad & f(x_1, x_2, \ldots, x_n) \\ \text{s.t.} \quad & g_1(x_1, x_2, \ldots, x_n) \leq / = / \geq b_1 \\ & g_2(x_1, x_2, \ldots, x_n) \leq / = / \geq b_2 \\ & \vdots \\ & g_n(x_1, x_2, \ldots, x_n) \leq / = / \geq b_n \end{aligned}$$

Note: $g_1, g_2 \ldots g_n$ are the constraints of the NLP. f is the objective function of the NLP.

An NLP with no constraints is called an unconstrained NLP.

An NLP with at least one constraint is called a constrained NLP.

☆ If either the objective function or (at least) one of the constraints becomes NL then the problem becomes an NLP instead of an LPP.

Def: The feasible region of an NLP is the set of points $(x_1, x_2, \ldots, x_n)$ that satisfy all the constraints. A point in the feasible region is called a feasible point and a point not in the feasible region is called an infeasible point.

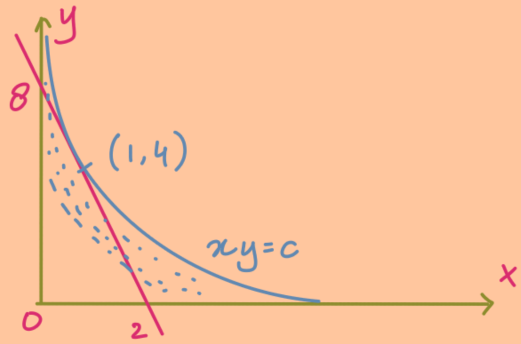
Def: For a maximization (minimization) problem any point $\bar{x}$ in the feasible region for which $f(\bar{x}) \geq (\leq) f(x)\ \forall\ x$ in the feasible region is called the optimal solution of the NLP.

Differences b/w NLP and LP

1) Feasible region of an NLP need not be a convex set.
2) In an NLP, even if the feasible region is a convex set, the optimal solution need not be at an extreme point.

Eg: max $z = xy$ s.t. $4x + y \leq 8$, $x, y \geq 0$

The slope of the curve at the point of x^* with the line $4x+y=8$ has to be the same as the slope of the line.



Slope of tangent of $xy = c \rightsquigarrow x\,dy + y\,dx = 0$

$\Rightarrow \dfrac{dy}{dx} = \dfrac{-y}{x} = -4$ ($y = 8-4x$)

So, the optimal point must satisfy both $y = 4x$ & $4x + y = 8$

$\Rightarrow y = 4,\ x = 1 \Rightarrow z^* = 4$

3) NLP need not satisfy additivity & proportionality assumption (But LP satisfies both of them)

Additivity Property: The contribution of any variable to the objective function is independent of the values of the other decision variables.

Ex: Consider xy — Increasing y by one unit will increase the objective function by x units. $\therefore$ xy does not satisfy additivity property.

Ex: Consider $5x+5y$ — Increasing y by one unit will increase the objective function by 5 units. Similarly, for x. Thus, $5x+5y$ satisfies additivity property.

Proportionality Property: If x is replaced by αx, then the contribution of x in the objective function also increases by a factor of α.

Ex: $x^{1/3}+y^{1/3}$ — Does not satisfy

Def: For any vector $x=(x_1, x_2, \dots, x_n)$, the Euclidean norm is defined as

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

p- norm is defined as

$$\|x\|_p = \begin{cases} (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{1/p}, & p<\infty \\ \max(x_1, x_2, \dots, x_n), & p=\infty \end{cases}$$

Def: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is a real valued function defined in some set $S \subseteq \mathbb{R}^n$. A point $x^* \in S$ is a local minima if $\exists$ $\epsilon > 0$ s.t. $f(x^*) \leq f(x)$, $\forall$ $x \in S$ s.t. $\|x - x^*\| < \epsilon$. For a local maxima,

$f(x^*) \geq f(x)$, rest of the conditions remain the same.

Def: A point that is a local minima or a local maxima is called a local or relative extrema.

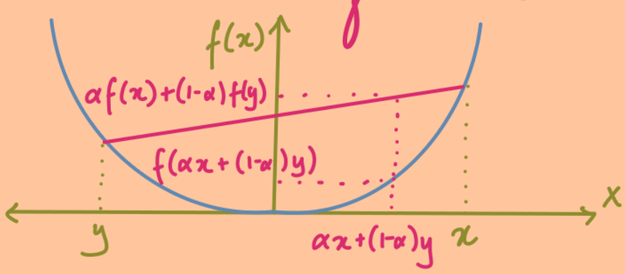
Def: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is a real valued function defined in some set $S \subseteq \mathbb{R}^n$. A point $x^* \in S$ is a global maxima (minima) if $f(x^*) \geq (\leq) f(x), \ \forall x \in S$

Def: Let S be any convex set in $\mathbb{R}^n$. Let f be a function defined on S. Then f is a convex function defined on a convex set S if $\forall x, y \in S$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y) \quad \forall \alpha \in [0,1]$$

→ The line joining the points $(x, f(x))$ & $(y, f(y))$ does not go below the curve in $[x, y]$ for convex functions.

Eg: $f(x) = x^2$ →



Def: Strictly convex function:

$$f(\alpha x + (1-\alpha) y) > \alpha f(x) + (1-\alpha) f(y)$$

Def: Concave function:

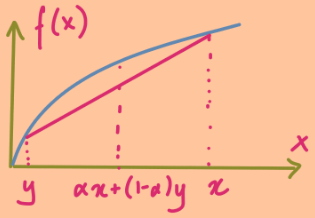
$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

S is still required to be convex though

Def: Strictly concave function:

$$f(\alpha x + (1-\alpha) y) < \alpha f(x) + (1-\alpha) f(y)$$

Ex: Concave function:



Notes: 1) Functions may be neither convex nor concave.

2) If $f: S \to \mathbb{R}$ is strictly concave (convex), then $-f$ will be a strictly convex (concave) function.

3) The sum of two convex (concave) functions is also a convex (concave) function.

4) A linear function is both a convex function as well as a concave function.

Theorem: Consider an NLP which is a maximization problem. Suppose the feasible region S is a convex set. If $f(x)$ is a concave function on S, then any local maxima for the NLP is an optimal solution (global maxima) of this NLP.

Pf: Suppose that $\exists\ \bar{x} \in S$ which is a local maxima but not a global maxima ... $\exists\ x$ in each nbd. of $\bar{x}$ s.t. $f(x) > f(\bar{x}) \Rightarrow \bar{x}$ is not a local maxima $\Rightarrow\Leftarrow$

☆ Similar theorem can be stated for a minimization NLP with a convex objective function in a convex set and local minimas.

Def: $f \in C^k$ if the k^{th} derivative exists and is continuous.

Def: A vector $d \in \mathbb{R}^n$, $d \neq 0$ is called a feasible direction at $x \in S$ if $\exists\ \alpha_0 > 0$ s.t. $x + \alpha d \in S \quad \forall\ \alpha \in [0, \alpha_0]$

Def: Let $f: \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$, $x = (x_1 \dots x_n) \in \mathbb{R}^n$, then $\underbrace{\nabla f(x)}_{\text{Gradient}} := \left[\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \dots \ \frac{\partial f}{\partial x_{n-1}} \ \frac{\partial f}{\partial x_n} \right] \in \mathbb{R}^n$

Def: The Hessian of $f: \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$ at $x \in \mathbb{R}^n$ is defined as

$$H(x) = \begin{bmatrix} \partial^2 f/\partial^2 x_1 & \partial^2 f/\partial x_1 \partial x_2 & \dots \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial^2 x_2 & \vdots \\ \vdots & \ddots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \dots & \partial^2 f/\partial^2 x_n \end{bmatrix}$$

Def: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a real-valued function let d be a feasible direction. Then the directional derivative of f in the direction of d is given by

$$\lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

and is denoted by $\nabla f(x)^T d$ or $d^T \nabla f(x)$

Taylor's Series Expansion

Let $f: \mathbb{R} \to \mathbb{R}$ & $f \in C^m$ on $[a, b]$. Define $h := b - a$, then

$$f(b) = f(a) + \frac{h}{1!} f'(a) + \frac{h^2}{2!} f^2(a) + \dots + \frac{h^m}{m!} f^m(a + \theta h)$$

for some $0 < \theta < 1$

☆ For a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, we have for $f \in C^2$,

$$f(x+h) = f(x) + \nabla f(x) h + \frac{1}{2} h^T H h \Big|_{x+\theta h}$$

for $0 < \theta < 1$, where H is the Hessian of f

Theorem: Let $f \in C^1$. Then f is a convex (concave) function over a convex set S iff

$$f(y) \geq (\leq)\ f(x) + \nabla f(x)^T (y-x) \quad \forall\, x, y \in S$$

Def: A point x is said to be an interior point of S if $\exists\ \epsilon > 0$ s.t.

$$B_x(\epsilon) := \{ y : \|y - x\| < \epsilon \} \subset S$$

☆ $\text{int } S := \{ x \in S : x$ is an interior point of $S \}$

Def: If $S = \text{int } S$, i.e., each point of S is an interior point, then S is called an open set.

Def: A quadratic form $f : \mathbb{R}^n \to \mathbb{R}$ is a function $f(x) = x^T Q x$ where Q is an $n \times n$ real matrix, i.e., $Q \in \mathbb{R}^{n \times n}$

Def: A quadratic form $x^T Q x$ with $Q = Q^T$ is said to be positive definite if

$$x^T Q x > 0 \quad \forall\ x \neq 0$$

Def: It is said to be positive semidefinite if $x^T Q x \geq 0 \quad \forall \; x$

Def: The quadratic form is said to be negative definite if $x^T Q x < 0 \quad \forall \; x \neq 0$
Similarly, negative semidefinite if
$$x^T Q x \leq 0 \quad \forall \; x$$

Def: The p^{th} order minor of an $m \times n$ matrix A with $p \leq \min(m, n)$ is the determinant of a $p \times p$ matrix obtained from A by deleting $m-p$ rows & $n-p$ columns.

Def: The p^{th} principal minor of an $n \times n$ matrix A is the determinant of a $p \times p$ matrix obtained from A by deleting $n-p$ rows & corresponding $n-p$ columns.

Def: The k^{th} leading principal minor of an $n \times n$ matrix A is the determinant of the $k \times k$ matrix obtained by deleting the last $n-k$ rows & the corresponding $n-k$ columns

Theorem: Let $f: \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$. Then f is convex (concave) over an open convex set S iff the Hessian matrix H of f is positive (negative) semidefinite throughout S.

Theorem: A quadratic form x^TQx, $Q = Q^T$ is positive definite iff the leading principal minors of Q have positive sign. [Sylvester's Criterion]

Theorem: A quadratic form x^TQx, $Q = Q^T$ is negative definite iff all the leading principal minors of order p of Q are of the same sign as $(-1)^p$.

Theorem: A quadratic form x^TQx, $Q = Q^T$ is positive semidefinite iff the principal minors of Q are non-negative.

Theorem: A quadratic form x^TQx, $Q = Q^T$ is negative semi-definite iff all the principal minors of order p of Q have the same sign as $(-1)^p$, if they are non-zero.

Theorem: Let S be an open convex set. Let $f \in C^2$. then f is a convex function on S iff for each $x \in S$ all principal minors of $H(f)$ are non-negative.

Pf: f is convex iff Hessian (H) of f is positive semi-definite throughout S.

Also, H of f is positive semidefinite throughout S iff all principal minors of H have non-negative determinant.

Theorem: Let S be an open convex set. Let $f \in C^2$, then f is a concave function on S iff for each $x \in S$ all non-zero principal minors of $H(f)$ of order p have the same sign as $(-1)^p$

Pf: Similar to the previous one.

Ex: Let $f(x_1, x_2) = -x_1^2 - x_1x_2 - 2x_2^2$ be defined over $\mathbb{R}^2$. Is f a convex / concave / NCC function?

$$\left.\begin{aligned} \partial f/\partial x_1 &= -2x_1 - x_2 \\ \partial f/\partial x_2 &= -x_1 - 4x_2 \end{aligned}\right\} \qquad H = \begin{bmatrix} -2 & -1 \\ -1 & -4 \end{bmatrix}$$

1st order principal minors: $-2, -4$

Same sign as $(-1)^1$

2nd order principal minor has determinant $(+7)$ of same sign as $(-1)^2$

$\therefore$ is a concave function.

Ex: Let $f(x_1, x_2) = x_1^2 - 3x_1x_2 + 2x_2^2$. Concave / Convex / Neither?

$$\begin{aligned} \partial f/\partial x_1 &= 2x_1 - 3x_2 \\ \partial f/\partial x_2 &= 4x_2 - 3x_1 \end{aligned} \qquad H = \begin{bmatrix} 2 & -3 \\ -3 & 4 \end{bmatrix}$$

1st order principal minors of H

have determinant 2, 4 > 0 ⟶ Not concave
2^{nd} order principal minor of H has determinant $8-9=-1<0$ ⟶ Not convex
$\therefore$ f is neither concave nor convex

Note: If f is a single variable function and $f''(x)$ exists for all x in an open convex set S, then $f(x)$ is a convex function on S iff $f''(x) \geq 0$ $\forall\, x \in S$. Similarly, $f(x)$ is a concave function on S iff $f''(x) \leq 0$ $\forall\, x \in S$

Ex: $f(x) = \sqrt{x}$, $x \in (0, \infty)$

$\hookrightarrow f'(x) = \dfrac{x^{-1/2}}{2} \longrightarrow f''(x) = -\dfrac{x^{-3/2}}{4} \leq 0$

$\therefore$ $\sqrt{x}$ is a concave function on $(0, \infty)$

Unconstrained Optimization with several variables

max/min $f(x_1, x_2, \dots, x_n)$, $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$

We would like to find local maxima/minima of the above function

First Order Necessary Condition

Let S be a subset of $\mathbb{R}^n$ and $f \in C'$ be a real valued function on S. x^* is a local minima/maxima of over f & x^* is an interior point of S, then $\nabla f(x^*) = 0$

Def: Points $\bar{x}$ for which $\nabla f(\bar{x}) = 0$ are called stationary points.

Second Order Necessary Condition

Let x^* be an interior point of $S \subseteq \mathbb{R}^n$. If x^* is a local minima (maxima) of function $f: S \to \mathbb{R}$, $f \in C^2$ then $\nabla f(x^*) = 0$ & $H(x^*)$ of f is positive (negative) semi-definite.

Second Order Sufficiency Condition

Let $f \in C^2$ be a function defined on a region in which the point x^* is an interior point. Suppose $\nabla f(x^*) = 0$ & $H(x^*)$ is positive (negative) definite, then x^* is a strict local minima (maxima) of f.

Ex: Consider the following unconstrained NLP defined for $x_1, x_2 \geq 0$

$$\max f(x_1, x_2) = x_1(70 - 4x_1) + x_2(150 - 15x_2) - 100 - 15x_1 - 15x_2$$

$$\nabla f(x) = 0 \longrightarrow \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = 0 = \begin{pmatrix} 70 - 8x_1 - 15, \\ 150 - 30x_2 - 15 \end{pmatrix}$$

$\Rightarrow x_1 = 55/8, \quad x_2 = 9/2 \longrightarrow$ Stationary Point

$$H = \begin{bmatrix} -8 & 0 \\ 0 & -30 \end{bmatrix} \longrightarrow$$ Need H to be negative definite

All leading principal matrices of order P should have the same sign as $(-1)^P$

1^{st} order leading principal minor $= -8 < 0$

2^{nd} order leading principal minor $= 240 > 0$

$\Rightarrow H(x^*)$ is negative definite.

$\Rightarrow (55/8, 9/2)$ is a strict local maxima of f

$\therefore$ H is negative definite $\forall (x_1, x_2) \in S$

$\Rightarrow f$ is strictly concave

$\Rightarrow (55/8, 9/2)$ is also the global maxima

Ex: Consider $f(x) = x^3$ on $\mathbb{R}$

$f'(x) = 3x^2 = 0$ at $x = 0$

$f''(x) = 6x = 0$ at $x = 0$

Thus, $x = 0$ satisfies the 2^{nd} Order Necessary Condition, but not the 2^{nd} Order Sufficiency Condition.

$x = 0$ fails to be a local maxima / minima

It is in fact, a saddle point

Def: If $|H(x^*)| \neq 0$ at a stationary point and if sufficiency conditions (2^{nd} Order) for maxima / minima are not satisfied, then the stationary point is called a saddle point.

Note: If $|H(x^*)| = 0$ at a stationary point x^*, then x^* may be a local maxima or local minima or a saddle point, and further investigation is required to conclude.

Theorem: If $f'(x^*) = 0$ and if the 1st non-vanishing (non-zero) derivative at x^* is

1) An odd-order derivative (for example $f^{(3)}(x)$), then x^* is neither a local maxima nor a local minima
2) Positive and an even-order derivative then x^* is a local minima
3) Negative and an even-order derivative then x^* is a local maxima

Eg: $f_1(x) = x^3$ and $f_2(x) = x^4$

$\hookrightarrow f_1^{(3)}(0) = 6 \neq 0 \Rightarrow 0$ is not an extrema

$f_1'(0) = f_2'(0) = 0$, $f_2^{(4)}(0) = 24 > 0 \Rightarrow 0$ is a minima

Eg: Let $f(x) = x(10-x)$, $0 < x < 10$. Find optimal solution for the following problem:

$\max f(x)$, $0 < x < 10$

$\hookrightarrow f'(x) = 10 - 2x \longrightarrow f'(x) = 0 \Rightarrow x = 5$

$\hookrightarrow f''(x) = -2 < 0 \longrightarrow f$ is concave in domain

$x = 5$ is a local + global maxima $\Leftarrow$

Constrained NLP

We shall start with the case when all constraints are equality constraints

Method of Lagrange Multipliers can be used to solve such NLPs.

Consider NLPs of the following type:

$$\max / \min \; f(x_1, x_2, \ldots, x_n)$$

s.t. $h_1(x_1, x_2, \dots, x_n) = 0$

$h_2(x_1, x_2, \dots, x_n) = 0$

$\vdots$

$h_m(x_1, x_2, \dots, x_n) = 0$

Def: A point $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ satisfying $h_i(x^*) = 0 \quad \forall \ i \in [m]$ is called a regular point of the constraints if the gradient vectors $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_m(x^*)$ are <u>linearly independent</u>.

$x_1, x_2, \dots, x_n$ are L.I. iff $\sum_{i \in [m]} \alpha_i x_i = 0 \Rightarrow \alpha_i = 0 \ \forall \ i \in [m]$

Def: The tangent space at a point x^* on the surface $S = \{x \in \mathbb{R}^n : \underbrace{h(x)=0}_{}\}$ is the set

$h = (h_1, h_2 \dots h_m)$

$1 \times m$ $\quad$ 0 vector

$$T(x^*) = \{y : \nabla h(x^*)^T y = 0\}$$

y is $\perp$ to each $\nabla h_i(x^*)$

0 vector, not scalar

Theorem: Let x^* be a local extremum point of f subject to the constraints $h(x) = 0$. Further, let x^* be a regular point of these constraints. Then, $\exists \ \lambda \in \mathbb{R}^m$ s.t.

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

First Order Necessary Conditions for x^* to be a local extrema

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

along with the constraints $h(x^*) = 0$

Notation: Lagrangian is denoted by L and defined as

$$L(x, \lambda) = f(x) + \lambda^T h(x)$$
$$\nabla_x L = \nabla_x f + \lambda^T \nabla h(x) = 0$$
$$\nabla_\lambda L = h(x) = 0$$ given constraints

Note: $\lambda_1, \lambda_2, \ldots, \lambda_m$ are called Lagrange Multipliers.

Second Order Necessary Conditions

- Let $f, h \in C^2$

Theorem: Suppose x^* is a local minima of f s.t. $h(x) = 0$ & that x^* is a regular point of these constraints. Then $\exists\ \lambda \in \mathbb{R}^m$ s.t. $\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$.
If we denote the tangent space by
$$M := \{ y : \nabla h(x^*)^T y = 0 \}$$
then $H_L := \nabla^2 f(x^*) + \lambda^T \nabla^2 h(x^*)$
(the Hessian of the Lagrangian)
is positive semidefinite on M, i.e.,
$$y^T H_L y \geq 0 \quad \forall \quad y \in M$$

Second Order Sufficiency Condition

Suppose x^* satisfies $h(x^*) = 0$
& $\lambda \in \mathbb{R}^m$ s.t. $\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$
Suppose that the matrix
$$H_L := \nabla^2 f(x^*) + \lambda^T \nabla^2 h(x^*)$$
is positive definite on M, i.e.,

$y^T H_L y > 0 \quad \forall \quad 0 \neq y \in M$. Then x^* is a strict local minima of $f(x)$ subject to $h(x) = 0$.

Note: Similar conditions can be written for x^* being a local maxima of $f(x)$ subject to $h(x) = 0$.

Note: max/min $f(x_1, \dots, x_n)$

s.t. $h_i(x_1, \dots, x_n) = 0 \quad \forall \; i \in [m]$

$L := f + \sum \lambda_i h_i$

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \sum_{i \in [m]} \lambda_i \frac{\partial h_i}{\partial x_1} = 0 \qquad \frac{\partial L}{\partial \lambda_1} = 0 \Rightarrow h_1 = 0$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} + \sum_{i \in [m]} \lambda_i \frac{\partial h_i}{\partial x_n} = 0 \qquad \frac{\partial L}{\partial \lambda_m} = 0 \Rightarrow h_m = 0$$

Let $(\bar{x}_1, \dots, \bar{x}_n, \bar{\lambda}_1, \dots, \bar{\lambda}_m)$ satisfy the $m+n$ equations above & all sufficiency conditions & maximize L.

Let $(x_1', \dots, x_n')$ be any point in the feasible region. Choose any $(\lambda_1', \dots, \lambda_m')$. Then

$$L(\bar{x}_i, \bar{\lambda}_j) \geq L(x_i', \lambda_j')$$

$$\Rightarrow f(\bar{x}) + \sum \bar{\lambda}_j \underbrace{h_j(\bar{x})}_{0 \text{ by def. of } \bar{x}} \geq f(x') + \sum \lambda_j' \underbrace{h_j(x')}_{0 \because x' \text{ is in the feasible region}}$$

$$\Rightarrow f(\bar{x}) \geq f(x') \longrightarrow \bar{x} \text{ maximizing } L \text{ also maximizes } f$$

Similarly, for a minimization problem.

Eg: $\max \; x_1 x_2 + x_2 x_3 + x_3 x_1$

s.t. $x_1 + x_2 + x_3 = 3, \quad (x_1, x_2, x_3) \in \mathbb{R}^3$

$L := (x_1 x_2 + x_2 x_3 + x_3 x_1) + \lambda(x_1+x_2+x_3-3)$

$$\left.\begin{aligned} \partial L/\partial x_1 &= x_2 + x_3 + \lambda = 0 \\ \partial L/\partial x_2 &= x_1 + x_3 + \lambda = 0 \\ \partial L/\partial x_3 &= x_2 + x_1 + \lambda = 0 \\ \partial L/\partial \lambda &= x_1 + x_2 + x_3 - 3 = 0 \end{aligned}\right\} \quad \begin{aligned} \lambda &= -2 \\ x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= 1 \end{aligned}$$

Check if the point is a regular point $\nabla h(x)$ has to be L.I.

$\nabla h = (1,1,1) \longrightarrow (\alpha)^T \cdot (1,1,1) = 0^3 \Rightarrow \alpha = 0$

$\Rightarrow \nabla h(x)$ is LI $\Rightarrow (1,1,1,-2)$ is a regular point

So, 1^{st} Order Necessary condition is satisfied.

→ Tangent plane is: $M := \{y : \nabla h(x)^T y = 0\}$

$\Rightarrow M = \{y : y_1 + y_2 + y_3 = 0\}$

And,

$H_L := \nabla^2 f + \lambda^T \nabla^2 h = \nabla(x_2+x_3,\ x_1+x_3,\ x_1+x_2) \underbrace{-2\nabla(1,\ 1,\ 1)}_{\text{is } 0}$

$\Rightarrow H_L = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Then, let $y \in M$ and we compute

$y^T H_L y = (y_1, y_2, y_3)\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$

$= (y_2+y_3,\ y_1+y_3,\ y_1+y_2)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = (-y_1, -y_2, -y_3)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$

$= -(y_1^2 + y_2^2 + y_3^2) < 0$

$\forall\ 0 \neq y \in M \Rightarrow H_L$ is negative definite

So, $x^* = (1,1,1)$ is a local maxima

Karush - Kuhn - Tucker Conditions

- Also called Kuhn - Tucker Conditions and KKT conditions.
- It is used in Constrained NLP with inequality conditions and is the heart of NLP problems.
- We consider problem of the form:

$$\max / \min \ f(x)$$
$$\text{s.t.} \quad h(x) = 0 \qquad (m \text{ constraints})$$
$$g(x) \leq 0 \qquad (p \text{ constraints})$$

Def: Let x^* be a point satisfying the constraints $h(x^*) = 0$, $g(x^*) \leq 0$ and let J be the set of indices j for which $g_j(x^*) = 0$. Then x^* is called a regular point of the given constraints if the gradient vectors $\nabla h_i(x^*)|_{i \in [m]}$ & $\nabla g_j(x^*)|_{j \in J}$ are linearly independent.

Def: Constraints among $g(x) \leq 0$ in which $g_j(x^*) = 0$ are called active / binding constraints.

KKT Conditions {Also the First Order Necessary Conditions}

Let x^* be a local minima for the problem $\min f(x)$ s.t. $h(x) = 0$, $g(x) \leq 0$ and suppose x^* is a regular point for the constraints. Then $\exists$ vectors $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ with $\underbrace{\mu \geq 0}_{\mu_i \geq 0 \ \forall \ i \in [p]}$ s.t. $\mu_i g_i(x^*) = 0 \ \forall \ i \in [p]$,

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0.$$

These new constraints, along with the given constraints are called KKT Constraints.

- For a local maxima, we want $\mu \leq 0$ instead of $\mu \geq 0$, rest of the conditions remain the same.

Second Order Necessary Conditions

Consider $\min f(x)$ s.t. $h(x) = 0$, $g(x) \leq 0$.

Suppose the functions $f, g, h \in C^2$ and that x^* is a regular point of the constraints. If x^* is a local minima for the problem, then $\exists\ \lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ s.t. $\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$,

$\mu_i g_i(x^*) = 0 \ \forall\ i \in [p]$ and

$$H_L := \nabla^2 L(x^*) = \nabla^2 f(x^*) + \lambda^T \nabla^2 h(x^*) + \mu^T \nabla^2 g(x^*)$$

(the Hessian of the Lagrangian) is positive semidefinite on the tangent space $M := \{ y : \nabla h(x^*)^T y = 0,\ \nabla g_j(x^*)^T y = 0\ \forall j \in J \}$ of the active constraints at x^*.

Second Order Sufficiency Conditions

Consider $\min f(x)$ s.t. $h(x) = 0$, $g(x) \leq 0$

Let $f, g, h \in C^2$. Sufficient conditions that a point x^* satisfying the constraints to be a strict local minima of f is that $\exists\ \lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ s.t. $\mu \geq 0$, $\mu^T g(x^*) = 0$, $\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$ and H of L at x^*,

$$H := \nabla^2 L(x^*) = \nabla^2 f(x^*) + \lambda^T \nabla^2 h(x^*) + \mu^T \nabla^2 g(x^*)$$
is positive definite on the tangent space
$$M := \{y : \nabla h(x^*)^T y = 0,\ \nabla g_j(x^*)^T y = 0\ \forall\ j \in J\}$$
of the active constraints at x^*.

☆ Similar constraints exist for a local maxima too.

Ex: Maximize $f(x) = x_1(30-x_1) + x_2(50-2x_2) - 3x_1 - 5x_2 - 10x_3$

s.t. $x_1 + x_2 \leq x_3, \quad x_3 \leq 17.25$

$$\mathcal{L} = x_1(30-x_1) + x_2(50-2x_2) - 3x_1 - 5x_2 - 10x_3 + \mu_1(x_1+x_2-x_3) + \mu_2(x_3 - 17.25)$$

$$\nabla \mathcal{L} = 0 \Rightarrow \frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial x_3} = 0, \quad \mu g(x^*) = 0$$

$\partial L/\partial x_1 = 30 - 2x_1 - 3 + \mu_1 = 0$

$\partial L/\partial x_2 = 50 - 4x_2 - 5 + \mu_1 = 0$

$\partial L/\partial x_3 = -10 - \mu_1 + \mu_2 = 0$

$\mu_1(x_1 + x_2 - x_3) = 0$

$\mu_2(x_3 - 17.25) = 0$

& $\mu_1, \mu_2 \leq 0$

[$\because$ it is a max. problem]

Cycle through the possibilities of μ_1, μ_2:

	$\mu_2 = 0$	< 0
$\mu_1 = 0$	1	2
$\mu_1 < 0$	3	4

(1) $\mu_1 = 0 = \mu_2 \rightsquigarrow -10 - \mu_1 + \mu_2 = -10 \neq 0 \Rightarrow\Leftarrow$

(2) $\mu_1 = 0 > \mu_2 \rightsquigarrow -10 - \mu_1 + \mu_2 = 0 \Rightarrow \mu_2 = 10 > 0 \Rightarrow\Leftarrow$

(4) $\mu_1 < 0 > \mu_2 \rightsquigarrow \mu_2 - \mu_1 = 10, \quad x_3 = 17.25,$

$45 + \mu_1 = 4x_2, \quad x_1 + x_2 = 17.25, \quad 27 + \mu_1 = 2x_1$

$\Rightarrow 4x_2 - 2x_1 = 18 \Rightarrow x_2 = 52.5/6 = 35/4$

$\Rightarrow \mu_1 = -10 \Rightarrow \mu_2 = 0 \Rightarrow\Leftarrow$

(3) $\mu_1 < 0 = \mu_2 \rightsquigarrow \mu_1 = -10 \Rightarrow x_1 + x_2 = x_3$

$\Rightarrow 2x_1 = 27 - 10 \Rightarrow x_1 = 17/2 = 8.5$

$\Rightarrow 4x_2 = 45 - 10 \Rightarrow x_2 = 35/4 = 8.75$

$\therefore$ Both the constraints are active $\Leftarrow \Rightarrow x_3 = 17.25$

Tangent plane: $M = \{y : \nabla(x_1+x_2-x_3)^T y = 0 = \nabla(x_3-17.25)^T y\}$

$\Rightarrow M = \{y : (1\ 1\ -1)^T (y_1\ y_2\ y_3) = 0 = (0\ 0\ 1)^T (y_1\ y_2\ y_3)\}$

$\Rightarrow M = \{y : y_1 + y_2 = y_3,\ y_3 = 0\} = \{y : y_1 + y_2 = 0 = y_3\}$

$\frac{\partial}{\partial x_1}\frac{\partial L}{\partial x_1},\ \frac{\partial}{\partial x_2}\frac{\partial L}{\partial x_1},\ \frac{\partial}{\partial x_3}\frac{\partial L}{\partial x_1}$

And Hessian of Lagrangian: $H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

Now, computing $y^T H y \ \forall\ 0 \neq y \in M$

$$y^T H y = (y_1\ y_2\ y_3) \begin{pmatrix} -2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = -2y_1^2 - 4y_2^2 < 0 \ \forall\ 0 \neq y \in M$$

Checking for regularity conditions; we require that both $\nabla(x_1+x_2-x_3) = (1, 1, -1)$ & $\nabla(x_3 - 17.25) = (0, 0, 1)$ are Linearly Independent ($\because$ both the constraints are active). So,

$\alpha_1(1, 1, -1) + \alpha_2(0, 0, 1) = (\alpha_1, \alpha_1, \alpha_2 - \alpha_1)$
$= (0, 0, 0) \Rightarrow \alpha_1 = 0 = \alpha_1 - \alpha_2 \Rightarrow \alpha_1 = 0 = \alpha_2$

Thus, $x^* = (8.5, 8.75, 17.25)$ is a regular point and satisfies all 1st & 2nd Order NASC [Necessary And Sufficient Conditions], $\therefore x^*$ is a local maxima.

$\because H$ satisfies Sylvester's Criterion, x^* is also a global maxima.

Note: If regularity conditions are not satisfied at an optimal point, the KKT conditions may fail at the optimal point.

Ex: max. $z = x_1$ s.t. $x_2 - (1-x_1)^3 \leq 0,\ x_1, x_2 \geq 0$

Now, $x_1 > 1 \Rightarrow x_2 < 0 \Rightarrow\Leftarrow$ So, x_1 cannot be >1. But $x_1 = 1 \Rightarrow x_2 \le 0$ and $x_2 \geq 0 \Rightarrow x_2 = 0$, So, $(1, 0)$ will maximize the objective function.

However, going by KKT:

$$L = x_1 + \mu_1[x_2 - (1-x_1)^3] + \mu_2[-x_1] + \mu_3[-x_2],$$

$$\mu_1, \mu_2, \mu_3 \le 0, \quad \mu_1(x_2 - (1-x_1)^3) = 0 = \mu_2 x_1 = \mu_3 x_2$$

$$\nabla L = (1 + 3\mu_1(1-x_1)^2 - \mu_2, \ \mu_1 - \mu_3)$$

at $(1, 0)$

$\Rightarrow \mu_2 = 0$ but $1 + 3\mu_1(1-1)^2 - \mu_2 = 1 \neq 0$

$\Rightarrow\Leftarrow$

Also, active constraints are $x_2 \le (1-x_1)^3$, $x_2 \geq 0$

Then, $\nabla(x_2 - (1-x_1)^3)$, $\nabla(-x_2)$ should be linearly independent at $(1, 0)$

$\nabla(x_2 - (1-x_1)^3)|_{(1,0)} = (0, 1)$, $\nabla(-x_2)|_{(1,0)} = (0, -1)$ are not linearly independent.

So, $x^* = (1, 0)$ is a local maxima without satisfying KKT conditions.

Quadratic Programming

The general quadratic programming problem can be expressed as

$$\min \ \tfrac{1}{2} x^T Q x + x^T c$$

$$\text{s.t.} \quad a_i^T x = b_i \quad \forall \ i \in [1, m]$$

$$a_j^T x \le b_j \quad \forall \ j \in [m+1, m+p]$$

where Q is symmetric and positive definite for minimization problems (and negative definite for maximization problems)

Ex: $\min \ 2x_1^2 + x_1 x_2 + x_2^2 - 12x_1 - 10x_2$

s.t. $x_1 + x_2 \leq 4, \quad x_1 \geq 0, \quad x_2 \geq 0$

$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad Q = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, \quad c = \begin{pmatrix} -12 \\ -10 \end{pmatrix}$

$a_1^T = (1, 1), \ b_1 = 4 \rightsquigarrow a_1^T x \leq b_1$

$a_2^T = (-1, 0), \ b_2 = 0 \rightsquigarrow a_2^T x \leq b_2$

$a_3^T = (0, -1), \ b_3 = 0 \rightsquigarrow a_3^T x \leq b_3$

☆ Constructing Q

$q_{ij} :=$ Co-efficient of $x_i x_j$ in f

$q_{ii} := 2 *$ Co-efficient of x_i^2 in f

Objective Function

But, more compactly, Q is the Hessian of the objective function f

☆ Since all the constraints are linear, we can assume the feasible region to be a convex set.

Also, $\because H \equiv Q$ is positive / negative definite, we'll have a convex / concave function on a convex set.

Thus, a local minima / maxima will also be a global minima / maxima.

Also, the constraints can be assumed to be linearly independent (redundant constraints can be removed), so that KKT conditions provide a solution / optimal point.

Ex: $\min. \ f = -x_1 - x_2 + \frac{1}{2} x_1^2 + x_2^2 - x_1 x_2$

s.t. $x_1 + x_2 \leq 3, \quad -2x_1 - 3x_2 \leq 6, \quad x_1, x_2 \geq 0$

$$Q = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad f = \tfrac{1}{2} x^T Q x + (-1\ -1)x$$

$$x_1 + x_2 \le 3$$
$$-2x_1 - 3x_2 \le -6$$
$$-x_1 \le 0, \quad -x_2 \le 0$$

$$L = -x_1 - x_2 + \tfrac{1}{2}x_1^2 + x_2^2 - x_1x_2 + \mu_1(x_1 + x_2 - 3) + \mu_2(-2x_1 - 3x_2 + 6) + \mu_3(-x_1) + \mu_4(-x_2),$$

$$\nabla L = 0 \longrightarrow \partial L/\partial x_1 = -1 + x_1 - x_2 + \mu_1 - 2\mu_2 - \mu_3 = 0^*$$
$$\partial L/\partial x_2 = -1 + 2x_2 - x_1 + \mu_1 - 3\mu_2 - \mu_4 = 0^*,$$

$$\mu_1(x_1 + x_2 - 3) = 0 = \mu_2(-2x_1 - 3x_2 + 6) = \mu_3(-x_1) = \mu_4(-x_2),$$

$$\mu_1, \mu_2, \mu_3, \mu_4 \geq 0,$$

$$x_1 + x_2 \le 3, \quad -2x_1 - 3x_2 \le -6, \quad -x_1 \le 0, \quad -x_2 \le 0$$

These are the KKT conditions.

$$x_1 + x_2 + s_1 = 3, \quad -2x_1 - 3x_2 + s_2 = -6^*,$$

Rewriting constraints using slack variables:

$$\{\mu_1(-s_1) = 0 = \mu_2(-s_2) = \mu_3(-x_1) = \mu_4(-x_2)\}$$

$$x_1, x_2, \mu_1, \mu_2, \mu_3, \mu_4, s_1, s_2 \geq 0$$

Complimentary Slackness conditions

(*) conditions do not have a basic variable, so we add an artificial variable to them and add them to get the objective function:

$$\min. \quad w = a_1 + a_2 + a_3$$

$$\text{s.t.} \quad \begin{aligned}
x_1 - x_2 + \mu_1 - 2\mu_2 - \mu_3 \qquad\qquad + a_1 \qquad &= 1 \\
-x_1 + 2x_2 + \mu_1 - 3\mu_2 \quad - \mu_4 \qquad + a_2 \qquad &= 1 \\
x_1 + x_2 \qquad\qquad + s_1 \qquad &= 3 \\
2x_1 + 3x_2 \qquad\qquad - s_2 \qquad + a_3 &= 6
\end{aligned}$$

$$\Rightarrow w = a_1 + a_2 + a_3 = 1 - x_1 + x_2 - \mu_1 + 2\mu_2 + \mu_3$$
$$+1 + x_1 - 2x_2 - \mu_1 + 3\mu_2 + \mu_4$$
$$+6 - 2x_1 - 3x_2 + s_2$$
$$= 8 - 2x_1 - 4x_2 - 2\mu_1 + 5\mu_2 + \mu_3 + \mu_4 + s_2$$

$$
\begin{aligned}
w + 2x_1 + 4x_2 + 2\mu_1 - 5\mu_2 - \mu_3 - \mu_4 \quad - s_2 &= 8 \\
x_1 - x_2 + \mu_1 - 2\mu_2 - \mu_3 \quad + a_1 &= 1 \\
-x_1 + 2x_2 + \mu_1 - 3\mu_2 \quad - \mu_4 \quad + a_2 &= 1 \\
x_1 + x_2 \quad + s_1 &= 3 \\
2x_1 + 3x_2 \quad - s_2 + a_3 &= 6
\end{aligned}
$$

Constructing the Simplex Table:

min	w	x_1	x_2	μ_1	μ_2	μ_3	μ_4	s_1	s_2	a_1	a_2	a_3	=
	1	2	4	2	-5	-1	-1		-1				8
a_1		1	-1	1	-2	-1				1			1
a_2		-1	$\boxed{2}$	1	-3		-1				1		1
s_1		1	1					1					3
a_3		2	3						-1			1	6
	1	4			1	-1	1		-1		-2		6
a_1		1/2		3/2	-7/2	-1	-1/2			1	1/2		3/2
x_2		-1/2	1	1/2	-3/2		-1/2				1/2		1/2
s_1		3/2		-1/2	3/2		1/2	1			-1/2		5/2
a_3		$\boxed{7/2}$		-3/2	9/2		3/2		-1		-3/2	1	9/2
	1			12/7	-20/7	-1	-5/7		1/7		-2/7	-8/7	6/7
a_1				12/7	-29/7	-1	-5/7		1/7	1	5/7	-1/7	6/7
x_2			1	2/7	-6/7		-2/7		-1/7		2/7	1/7	8/7
s_1				1/7	-3/7		-1/7	1	3/7		1/7	-3/7	4/7
x_1				-3/7	9/7		3/7		-2/7		3/7	2/7	9/7

⋮ Final Simplex Table Below ⋮

	1									-1	-1	-1	0
μ_1				1	-12/5	-3/5	-2/5	-1/5		3/5	2/5		2/5
x_2			1		-1/5	1/5	-1/5	2/5		-1/5	1/5		6/5
s_2					-1/5	1/5	-1/5	12/5	1	-1/5	1/5	-1	6/5
x_1		1			1/5	-1/5	1/5	3/5		1/5	-1/5		9/5

This is the Wolfe's Method, a variant of Simplex method to solve Quadratic Programming.

The only variation is while choosing the entering variable. We do not allow two variables to be basic if their being basic simultaneously leads to a violation of some complimentary slackness condition.

For example, in the highlighted simplex table, normally for a minimization problem, we would have chosen μ_1, but we cannot choose it as making it a basic variable along with s_1 would have led to violation of $\mu_1 s_1 = 0$. So, we chose the next variable (s_2 here) as the entering variable.

We keep running the algorithm until the objective function (w) becomes 0.

Fin