C1 Introduction to Transport

Question 1

Consider the weighted graph $G = (V, E)$ illustrated in the figure below.

a. Calculate the degree of every node. How about indegrees and outdegrees?

b. Provide the node adjacency matrix for the graph, and a forward star representation that would include link weights.

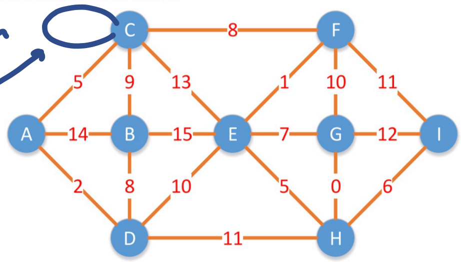
connectivity representation

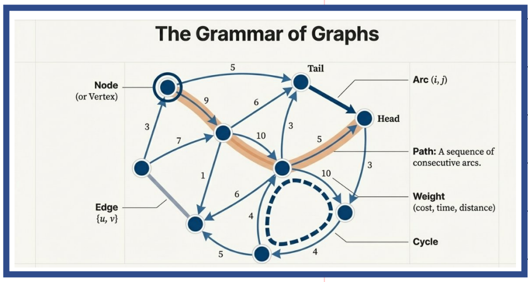
c. Can an adjacency matrix by used to calculate the degrees of a node i? Provide a formula.

d. How would you represent a buckle added to node C in the adjacency matrix?

e. Obtain a degree distribution for this table.

info: (d) a buckle is when a node is connected to itself with one edge.





Solution 1a

degree is no. of edges connected

indegree and out degrees are for directed graph.

This is an undirected graph – as such, nodes only have degrees.

Node	Degree
A	3
B	4
C	4
D	4
E	6
F	4
G	4
H	4
I	3

Solution 1b

two different way to represent graph connectivity

The node adjacency matrix for the graph is as follows: to

from

	A	B	C	D	E	F	G	H	I
A	0	1	1	1	0	0	0	0	0
B	1	0	1	1	1	0	0	0	0
C	1	1	0	0	1	1	0	0	0
D	1	1	0	0	1	0	0	1	0
E	0	1	1	1	0	1	1	1	0
F	0	0	1	0	1	0	1	0	1
G	0	0	0	0	1	1	0	1	1
H	0	0	0	1	1	0	1	0	1
I	0	0	0	0	0	1	1	1	0

show weight

A forward star representation for the graph is as follows:

Node
A
B
C
D
E
F
G
H
I

Node1	Node2	Cost
A	C	5
C	A	5
A	B	14
B	A	14
A	D	2
D	A	2
C	B	9
B	C	9
C	E	13
E	C	13
C	F	8
F	C	8
B	E	15
E	B	15
D	B	8

(continued)

Node1	Node2	Cost
B	D	8
D	E	10
E	D	10
D	H	11
H	D	11
E	F	1
F	E	1
E	H	5
H	E	5
E	G	7
G	E	7
F	G	10
G	F	10
G	H	0
H	G	0

(continued)

Node1	Node2	Cost
F	I	11
I	F	11
G	I	12
I	G	12
H	I	6
I	H	6

- if this is directed graph there will not be A→C and C→A so overall the table would be half the size.

Solution 1c

For every node $\mathrm{i} \in \mathrm{V}$, the degree will be calculated from the following function, where $\mathrm{k_{i,j}}$ is the value of the adjacency matrix in position $\mathrm{i, j}$:

$$\sum_{j \in V} k_{i,j}$$

Solution 1d

Since this graph is undirected, the adjacency matrix is transposable – a value of 1 on the cell C,C would not capture the bidirectional nature of the buckle. Instead, the convention is to use a value of 2.

this is not weight! this just show undirected graph $C \to C$ should count as two!

Solution 1e

The degree distribution for the graph is as follows:

Degree	Occurrence	Probability
1	0	0
2	0	0
3	2	0.222
4	6	0.667
5	0	0
6	1	0.111

Question 2
Suppose that every node in a directed graph has a nonzero indegree. Show that the graph must contain a directed cycle. Hint: use an example graph in your response.

Solution 2

Let $\mathrm{i_1}$ be any arbitrary node in the graph. Since $\mathrm{i_1}$ has a positive indegree, there exists some node $\mathrm{i_2}$ such that $(\mathrm{i_2}, \mathrm{i_1}) \in \mathrm{A}$, where A is the set of all edges.

Again, since $\mathrm{i_2}$ has a positive indegree, there exists some node $(\mathrm{i_3}, \mathrm{i_2}) \in \mathrm{A}$. If $\mathrm{i_3} = \mathrm{i_1}$, a directed cycle has just been discovered in the graph; otherwise we continue this process until we visit a node twice.

Notice that $\mathrm{i_1}$ will eventually be reached since the graph has a finite number of nodes. In this case, we have discovered a directed cycle in the graph.

C2. Network Algorithm

Network Algorithm.

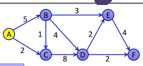
Shortest Path Problem

Clustering Problem.
- to identify group with similar characteristic within a collection of element.

Dijkstra's Algorithm ← (no negative weight arc!) → A* Algorithm.

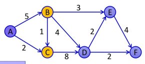
- Modification of Dijkstra's (used for larger network)
- Differences: Instead of checking all nodes, it select next node based on least estimated total cost. (a heuristic approach)
- point-to-point path finding algo.

Tabular Implementation
Step 1



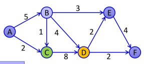
Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A						
C	∞ A						
D	∞ A						
E	∞ A						
F	∞ A						

Tabular Implementation
Step 2



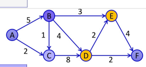
Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A	0+5=5 A					
C	∞ A	0+2=2 A					
D	∞ A	∞ A					
E	∞ A	∞ A					
F	∞ A	∞ A					

Tabular Implementation
Step 3



Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A	5 A	5 A				
C	∞ A	2 A					
D	∞ A	∞ A	2+8=10 C				
E	∞ A	∞ A	∞ A				
F	∞ A	∞ A	∞ A				

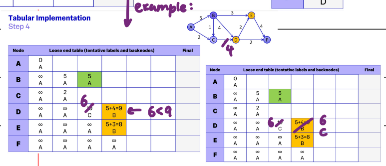
Tabular Implementation
Step 4



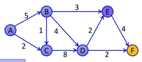
Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A	5 A	5 A				
C	∞ A	2 A					
D	∞ A	∞ A	10 C	5+4=9 B			
E	∞ A	∞ A	∞ A	5+3=8 B			
F	∞ A	∞ A	∞ A	∞ A			

← if the newly expanded is larger than the previous cell, carry the previous forward:

example:

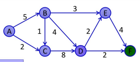


Tabular Implementation
Step 5



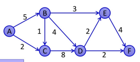
Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A	5 A	5 A				
C	∞ A	2 A					
D	∞ A	∞ A	10 C	9 B	9 B		
E	∞ A	∞ A	∞ A	8 B			
F	∞ A	∞ A	∞ A	∞ A	8+4=12 E		

Tabular Implementation
Step 6



Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						
B	∞ A	5 A	5 A				
C	∞ A	2 A					
D	∞ A	∞ A	10 C	9 B	9 B		
E	∞ A	∞ A	∞ A	8 B			
F	∞ A	∞ A	∞ A	∞ A	12 E	9+2=11 D	

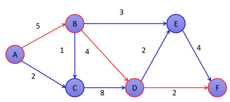
Tabular Implementation
Step 7



Node	Loose end table (tentative labels and backnodes)						Final
A	0 A						0 A
B	∞ A	5 A	5 A				5 A
C	∞ A	2 A					2 A
D	∞ A	∞ A	10 C	9 B	9 B		9 B
E	∞ A	∞ A	∞ A	8 B			8 B
F	∞ A	∞ A	∞ A	∞ A	12 E	11 D	11 D

Tabular Implementation
Step 8

Node	Final
A	0 A
B	5 A
C	2 A
D	9 B
E	8 B
F	11 D



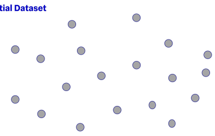
Final path cost (label): 11

Route: F ← D ← B ← A

(Heuristic)

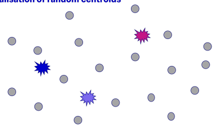
K-mean Clustering:

Step 0 - Initial Dataset



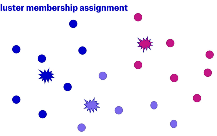
We are given an initial set of elements that we need to cluster. In this case, we need to only perform geospatial clustering, and will focus on two properties only, their x and y coordinates.

Step 1 - Initialisation of random centroids



The k-means algorithm requires a known value of k when initialised. If we are not sure about the number of clusters that we want, we can try different values. At initialisation, the algorithm will randomly pick a starting centroid for the first iteration.

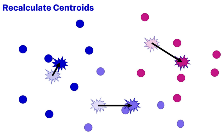
Step 2a - Cluster membership assignment



Using the distance metric that we have chosen, we allocate every element to the nearest cluster centroid. We do pick the centroid m that every element i is assigned to, using:

$$d_{im} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2}$$

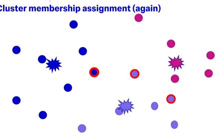
Step 3a – Recalculate Centroids



We now recalculate the centroids $\vec{\mu}$ for every individual cluster c, using:

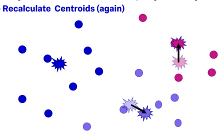
$$\vec{\mu}(c) = \frac{1}{|c|}\sum_{\vec{x} \in c} \vec{x}$$

Step 2b – Cluster membership assignment (again)



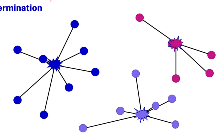
We perform, again, a new round of cluster membership assignments using the new centroids.

Step 3b – Recalculate Centroids (again)



Unless the new cluster membership assignments coincide perfectly with the ones determined previously, we proceed with a new round of centroid calculations

Step 4 – Termination



- We terminate our iterations once no more changes can be performed in our set of clusters
- The overall approach seeks to minimise the inertia of the assignment, or in other words, the sum of distances between the elements and their corresponding cluster centroids.

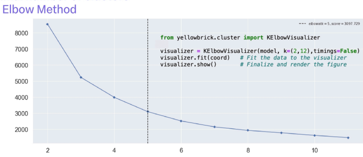
Limitation of K-Means

- **Initialization:** The algorithm requires a **known value of *k*** (number of clusters) when initialised. You often have to guess or use the "Elbow Method" to find the right number.
- **Randomness:** At initialisation, the algorithm randomly picks starting centroids. Different random starts can lead to different clustering results.
- **Operational Limits:** Standard clustering (like K-means based on distance) does not automatically take into account operational costs (like fixed warehouse costs vs transportation costs) unless explicitly modelled.

so how many k? (optimal)

Elbow Method
- if optimal doesn't exist

Optimal number of clusters
The Elbow Method



- The elbow method is a generic technique that is used to determine the optimal number of clusters in a dataset.
- After plotting the inertia values for a range of k values, we choose the value that lies in the inflection point of the curve (i.e. the "elbow")

(how much is enough?)

Operational Modelling
- if optimal exist.

Optimal number of clusters
Operational Modelling

Clusters	Distances	Warehousing		Transportation		Total
3	5231	1500		4184.8		5684.8
4	3983	2000		3186.4		5186.4
5	3097	2500	+	2477.6	=	4977.6
6	2518	3000		2014.4		5014.4
7	2147	3500		1717.6		5217.6
8	1937	4000		1549.6		5549.6

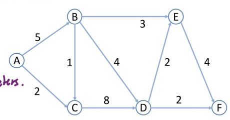
- While the elbow method can be conveniently applied to any dataset, it does not take into account any further operational aspects of the problem.
- In the above example we assume that every warehouse has a fixed operational cost of 500 units per week.
- We assume that the cost of transportation is 0.75 per distance unit.
- Using the above, we can determine that costs are minimised for 5 clusters.

Shortest Path Problem Math Formulation:

Shortest Path Problems – Abstract Formulation

N	*Set of vertices*
E	*Set of links in the network*
x_{ij}	*Flow on arc (i,j)* Decision Variable.
C_{ij}	*Cost of transport on arc (i,j)*
s	*Source node (part of N)*
t	*Terminal node (part of N)*

Parameters.



minimize $\sum_{(i,j)\in E} C_{ij}x_{ij}$ ← **Objective** *Same as before*

subject to $\sum_{j:(i,j)\in E} x_{ij} - \sum_{k:(k,i)\in E} x_{ji} = D_i \;\; \forall\, i \in N$ ← **Constraint 1** *Same as before*

$$D_i = \begin{cases} 1, & \text{for } i = s \\ -1 & \text{for } i = t \\ 0 & \text{otherwise} \end{cases}$$ ← **Constraint 2** *Defines endpoints*

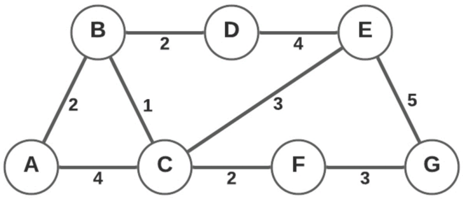
$x_{ij} \in \{0,1\} \quad \forall\, (i,j) \in E$ ← **Constraint 3** *Boolean flows*

- Minor modifications were on constraints.
- These ensure that flows indicate a single route.
- Convenient approach, but not computationally efficient

(derived from c3: min cost pro. with transit)
add/modify: ONLY single sink and single source node, boolean decision variable.

Question 2 **(how to use Dijkstra's to find closeness centrallity)**

The figure below illustrates a network graph. b. Use Dijkstra's algorithm to calculate the closeness centrality for vertex D.



Solution 2

start from D →

Node	Loose End Table A to G							Final
A	∞	∞	4 *B*	4 *B*	–			4 *B*
B	∞	2 *D*	–					2 *D*
C	∞	∞	3 *D*	–				3 *D*
D	0 *D*	–						0 *D*
E	∞	4 *D*	4 *D*	4 *D*	4 *D*	–		4 *D*
F	∞	∞	∞	5 *C*	5 *C*	5 *C*	–	5 *C*
G	∞	∞	∞	∞	∞	9 *E*	8 *F*	8 *F*

D→B→A=4

D→B→C→A=7

Node	Loose End Table A to G							Final
A	∞	∞	4 *B*	4 *B*	–			4 *B*
B	∞	2 *D*	–					2 *D*
C	∞	∞	3 *D*	–				3 *D*
D	0 *D*	–						0 *D*
E	∞	4 *D*	4 *D*	4 *D*	4 *D*	–		4 *D*
F	∞	∞	∞	5 *C*	5 *C*	5 *C*	–	5 *C*
G	∞	∞	∞	∞	∞	9 *E*	8 *F*	8 *F*

at this stage both are the same, expand any of it: A

but since A only A→B, A→C and B and C already expanded, A is ended.

Node	Loose End Table A to G							Final
A	∞	∞	4 *B*	4 *B*	–			4 *B*
B	∞	2 *D*	–					2 *D*
C	∞	∞	3 *D*	–				3 *D*
D	0 *D*	–						0 *D*
E	∞	4 *D*	4 *D*	4 *D*	4 *D*	–		4 *D*
F	∞	∞	∞	5 *C*	5 *C*	5 *C*	–	5 *C*
G	∞	∞	∞	∞	∞	9 *E*	8 *F*	8 *F*

when C is expanded, we need to check if the new path can replace the previous shortest path.

The average distance from D to the other nodes in the network is 4.333 and therefore the closeness centrality of node D is 0.23.

$$\frac{4+2+3+4+5+8}{6} = 4.333$$

closeness centrallity = 1 ÷ average distance

= 1 ÷ 4.333

= 0.231

In graph theory, centrality is defined as a measure of the importance of a node in the graph.

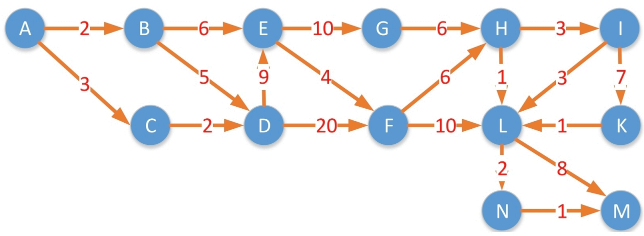
- Degree centrality: Number of links upon a node. **(same as finding degree table in c1)**
- Closeness centrality: Inverse of average path length from a specific node to all other nodes in the network.
- Betweenness centrality: Number of times a node is included among the set of shortest paths in the network.
- Eigenvector centrality: A node is more important if it connects to more important nodes (similar to PageRank, used by Google)

Question 3

a. Using Dijkstra's algorithm, determine the shortest path between nodes A and M for the graph below.

b. Describe the steps involved in the algorithm that would find the shortest path between an origin and destination pair, subject to the additional condition that the path must visit a specified node.

c. For the graph shown below, find a path between A and M that also visits node G.



Solution 3a

We set up a loose end table with all the nodes and traverse the network using the Dijkstra’s algorithm.

Node	**Loose End Table**														**Final**
A	0 A	.													0 A
B	∞ A	0+2=2 A	.												2 A
C	∞ A	0+3=3 A	3 A	.											3 A
D	∞ A	∞ A	2+5=7 B	3+2=5 C	.										5 C
E	∞ A	∞ A	2+6=8 B	8 B	8 B	.									8 B
F	∞ A	∞ A	∞ A	∞ A	5+20=25 D	8+4=12 E	.								12 E
G	∞ A	∞ A	∞ A	∞ A	∞ A	8+10=18 E	18 E	.							18 E
H	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	12+6=18 F	18 F	.						18 F
I	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	18+3=21 H	21 H	.				21 H
K	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	21+7=28 I	28 I	28 I	.	28 I
L	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	12+10=2 2 F	22 F	18+1=19 H	.					19 H
M	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	19+8=27 L	27 L	21+1=22 N	.		22 N
N	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	∞ A	19+2=21 L	21 L	.			21 L

According to the final column, the total cost of travelling from A to M is 22. If we backtrack from M using the predecessor labels in the final column, the optimal path is: M ← N ← L ← H ← F ← E ← B ← A

Solution 3b

Many approaches could be used to solve this problem – an acceptable solution would suggest that we specify an algorithm that finds the shortest path from A to G and another from G to M. The final path would be a combination of the two paths.

Solution 3c

Node	Loose End Table A to G						Final
A	0 A						0 A
B	∞ A	$0+2=2$ A					2 A
C	∞ A	$0+3=3$ A	3 A				3 A
D	∞ A	∞ A	$2+5=7$ B	$3+2=5$ C			5 C
E	∞ A	∞ A	$2+6=8$ B	8 B	8 B		**8** **B**
F	∞ A	∞ A	∞ A	∞ A	$5+20=25$ D	$8+4=12$ E	**12** **E**
G	∞ A	∞ A	∞ A	∞ A	∞ A	$8+10=18$ E	**18** **E**

Node	Loose End Table G to M						Final
G	0 G						**0** **G**
H	∞ G	$0+6=6$ G					**6** **G**
I	∞ G	∞ G	$6+3=9$ H	9 H	9 H		**9** **H**
K	∞ G	∞ G	∞ G	∞ G	∞ G	$9+7=16$ I	**16** **I**
L	∞ G	∞ G	$6+1=7$ H				**7** **H**
M	∞ G	∞ G	∞ G	$7+8=15$ L	$9+1=10$ N	10 N	**10** **N**
N	∞ G	∞ G	∞ G	$7+2=9$ L			**9** **L**

From A to G: A → B → E → G, cost: 18
From G to M: G → H → L → N → M, cost: 10
Path between A and M, that also visits node G: A→ B→ E→ G → H → L → N → M, total cost: 28

Question 4

a. Apply Dijkstra's algorithm to obtain the shortest path between nodes A and D at the weighted digraph provided in Figure 2.
b. Comment on your findings and suggest an additional step in your approach that would address any issues that you have encountered.
c. Undertake your own research to identify an algorithm that is capable of determining paths in graphs with negative edge weights.

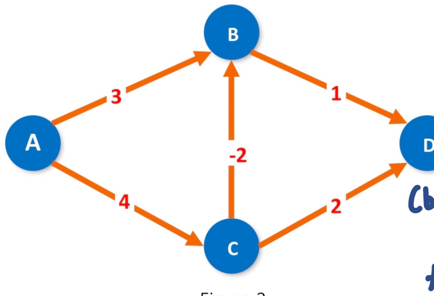


Figure 2

(b) suggested step:
if found a shorter path that already visited, marked as unvisited again and add back to processing queue

(a)

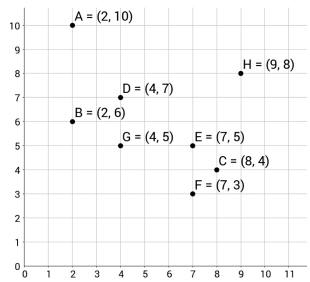
A	0 A	-			
B	∞ A	3 A	-		
C	∞ A	4 A	4 A	-	
D	∞ A	∞ A	4 B	4 B	-

expand C as it ties with D
C→D=6 < 4 (does not replace)

can't revisit B as already close case.

Question 6 (K-Mean Question)

A contractor has 8 project sites with the coordinates and Euclidian distances given in the table below. The company has decided that 2 distribution centres will be established to supply these sites with materials.



Site	X	Y	d(A)	d(B)	d(C)	d(D)	d(E)	d(F)	d(G)	d(H)
A	2	10	0.00	4.00	8.49	3.61	7.07	8.60	5.39	7.28
B	2	6		0.00	6.32	2.24	5.10	5.83	2.24	7.28
C	8	4			0.00	5.00	1.41	1.41	4.12	4.12
D	4	7				0.00	3.61	5.00	2.00	5.10
E	7	5					0.00	2.00	3.00	3.61
F	7	3						0.00	3.61	5.39
G	4	5							0.00	5.83
H	9	8								0.00

a. Use a pen-and-paper (and calculator) approach to follow the steps involved in the k-means clustering algorithm to obtain two clusters for the above points. You can start with points A and C as initial guess for the centre points.

Solution 6a

Step 1:

Site	d(P1)	d(P2)
A	0.00	8.49
B	4.00	6.32
C	8.49	0.00
D	3.61	5.00
E	7.07	1.41
F	8.60	1.41
G	5.39	4.12
H	7.28	4.12

Cluster 1: {A, B, D}
Cluster 2: {C, E, F, G, H}

New centre points:

$$\mathrm{P1} = \left(\frac{2+2+4}{3}, \frac{10+6+7}{3}\right) = (2.67, 7.67)$$

$$\mathrm{P2} = \left(\frac{8+7+7+4+9}{5}, \frac{4+5+3+5+8}{5}\right) = (7.00, 5.00)$$

Step 2:

Site	d(P1)	d(P2)
A	2.43	7.07
B	1.80	5.10
C	6.47	1.41
D	1.49	3.61
E	5.09	0.00
F	6.37	2.00
G	2.98	3.00
H	6.34	3.61

Cluster 1: {A, B, D, G}
Cluster 2: {C, E, F, H}

New centre points:

$$\mathrm{P1} = \left(\frac{2+2+4+4}{4}, \frac{10+6+7+5}{4}\right) = (3.00, 7.00)$$

$$\mathrm{P2} = \left(\frac{8+7+7+9}{4}, \frac{4+5+3+8}{4}\right) = (7.75, 5.00)$$

Step 3:

Site	d(P1)	d(P2)
A	3.16	7.62
B	1.41	5.84
C	5.83	1.03
D	1.00	4.25
E	4.47	0.75
F	5.66	2.14
G	2.24	3.75
H	6.08	3.25

The cluster partition has not changed with respect to the previous step. Therefore, the optimal clusters are:

Cluster 1: {A, B, D, G}
Cluster 2: {C, E, F, H}

c3 Mathematical Models (Minimise Cost Flow Problem)

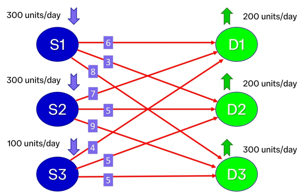
Class of Model:

Linear Programs (LP) *(objective/constraints are linear formulas, variables are Real)*	$minimize\ 2x + y$ x, y: Real
Integer Programs (IP) *(all decision variables are integers, formulas are linear)*	$maximize\ x + 5y$ $x \in [1,2, \ldots 5], y \in [0.1]$
Mixed Integer Linear Programs (MILP) *(linear programs, mix of integer and linear decision variables)*	$minimize\ 2x - 5y$ x: Real, $y \in [5,6, \ldots 10]$
Quadratic Programs (QP) *(quadratic objective function, linear constraints)*	$minimize\ x + y^2$ x, y: Real

Mathematical Models

What is it?

1. Minimum cost problem / Maximise Profit (Supply Demand)



Mathematical Models
Explicit Form

$x_{S_1D_1} \ldots x_{S_3D_3}$ Flow quantity	
minimize	$6x_{S_1D_1}+3x_{S_1D_2}+8x_{S_1D_3}+7x_{S_2D_1}+5x_{S_2D_2}+9x_{S_2D_3}+4x_{S_3D_1}+5x_{S_3D_2}+5x_{S_3D_3}$
subject to	$300 = x_{S_1D_1}+x_{S_1D_2}+x_{S_1D_3}$ $200 = x_{S_1D_1}+x_{S_2D_1}+x_{S_3D_1}$
	$300 = x_{S_2D_1}+x_{S_2D_2}+x_{S_2D_3}$ $200 = x_{S_1D_2}+x_{S_2D_2}+x_{S_3D_2}$
	$100 = x_{S_3D_1}+x_{S_3D_2}+x_{S_3D_3}$ $300 = x_{S_1D_3}+x_{S_2D_3}+x_{S_3D_3}$
	$x_{S_1D_1}, x_{S_1D_2}, x_{S_1D_3}, x_{S_2D_1}, x_{S_2D_2}, x_{S_2D_3}, x_{S_3D_1}, x_{S_3D_2}, x_{S_3D_3} \geq 0$

Mathematical Models (min cost pro. without transit) * known as a transportation prob!
Indexed-Symbolic Form

S	Set of supply nodes in the network	We define sets
D	Set of demand nodes in the network	
x_{ij}	Flow on link (i,j)	Decision Variable.
C_{ij}	Cost of transport on link (i,j)	Variable and parameter definitions now use component sets
A_i	Cargo supply in node i	
B_i	Cargo demand in node i	Parameter
minimize	$\sum_{i \in S} \sum_{j \in D} C_{ij} x_{ij}$	Objective Function
subject to	$\sum_{j \in D} x_{ij} = A_i \quad \forall\, i \in S$ $\sum_{i \in S} x_{ij} = B_j \quad \forall\, j \in D$	Constraints. One of the constraints usually can have $\leq$ instead of $=$
	$0 \leq x_{ij} \leq L_{ij} \quad \forall i \in S, j \in D$	

(L_{ij} is if link/arc has capacity too)

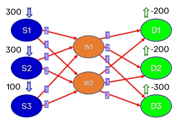
Decision Variable can be:
- linear (continuous)
- Integer (discrete)
- boolean (0 or 1)

Objective Function can be:
- single: $\max f(x)$ or $\min f(x)$
  multiple: $\max f(x)$ and $\min g(x)$
  (can be written as: $\max\{Af(x) - Bg(x)\}$)
- Maximin or Minimax: $\operatorname{Min}(\operatorname{Max} f(x))$
  (can be written as: $\operatorname{Min}_z (f(x) - z)$)

Transhipment (min cost pro. with transit)
Symbolic formulation

- The modified representation of the network allows us to use a significantly simpler model formulation.
- A single constraint now captures production, consumption and transhipment

N	Set of nodes	
E	Set of transportation links	
x_{ij}	Flow of goods between i, j	Symbol definitions
C_{ij}	Cost of transport on arc i, j	
D_i	Cargo supply/demand on vertex i	
minimize	$\sum_{(i,j) \in E} C_{ij} x_{ij}$	Objective function
subject to	$\sum_{j:(i,j) \in E} x_{ij} - \sum_{j:(\cancel{k,i})\,(j,i) \in E} x_{\cancel{ki}\,ji} = D_i \quad \forall\, i \in N$	Constraint 1 Flow Conservation
	$0 \leq x_{ij} \leq L_{ij} \quad \forall i, j \in E$	Constraint 2 Non-Negativity



* known as a transhipment prob.

19/10/2024

in this example we only transhipping one type of good.

out − in = capacity ✓

capacity of supply +ve
capacity of demand −ve
transhipment node usually 0

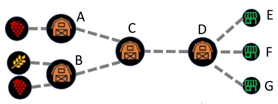
3. For multicommodity flow

N	Set of vertices
E	Set of arcs
P	Set of commodities
x_{ij}^c	Flow of commodity c on arc (i,j)
C_{ij}	Cost of transport on arc (i,j)
D_i^c	Production of commodity c on vertex i
L_{ij}	Capacity limit on arc (i,j)
minimize	$\sum_{c \in P} \sum_{(i,j) \in E} C_{ij} x_{ij}^c$
subject to	$\sum_{j:(i,j) \in E} x_{ij}^c - \sum_{j:(j,i) \in E} x_{ji}^c = D_i^c \quad \forall\, i \in N, c \in P$
	$\sum_{c \in P} x_{ij}^c \leq L_{ij} \quad \forall\, i,j \in E$
	$0 \leq x_{ij}^c \quad \forall i,j \in E, c \in P$

- An variable L_{ij} has been added to represent the capacity limit on each arc.
- A new constraint imposes an upper limit on the sum of flows within a link.

we minimise total cost (this inc. summing all c)

we track flow cons. seperate (eg. grape in = grape out)



but for capacity limit, each link limit doesn't care what commodity it carries but so sum of all the c must adhere the capacity limit.

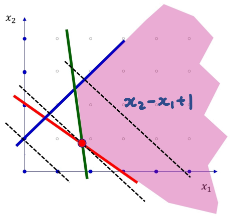
(eg. a train doesn't care it carries apple or orange but it can't exceed its capacity)

How to solve?

Method 1:

Constructing an LP graph

$$\min x_1 + x_2$$
$$2x_1 + 3x_2 \geq 6$$
$$-x_1 + 2x_2 \leq 2$$
$$5x_1 + x_2 \geq 10$$

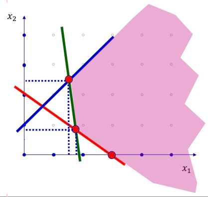


min $x_1 + x_2$
plot $x_1 + x_2 = k$ (let $k = 1$):
$x_2 = -x_1 + 1$
↑ m   ↑ c

Method 2:

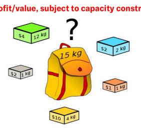
Constructing an LP graph

Corner Point	Value
$x_1 = 3.0, x_2 = 0.0$	$3 + 0 = 3$
$x_1 = 1.6, x_2 = 1.8$	$1.6 + 1.8 = 3.4$
$x_1 = 1.8, x_2 = 0.7$	$1.8 + 0.7 = 2.6$



2. Knapsack Problem
(Maximise Profit subject to capacity constraint)

Maximise total profit/value, subject to capacity constraints.



The Knapsack Problem
Explicit Model

x_1, x_2, x_3, x_4	Quantity of each type of item	**Decision Variables**
maximise	$1000x_1 + 7000x_2 + 3000x_3 + 250x_4$	**Objective Function**
subject to		**Constraints**
	$50x_1 + 450x_2 + 100x_3 + 20x_4 \leq 16200$	*Weight capacity*
	$0.4x_1 + x_2 + 2x_3 + 0.1x_4 \leq 32$	*Volume capacity*
	$x_1, x_2, x_3, x_4 \geq 0$	*Non-Negativity*
	$x_1, x_2, x_3, x_4 \in \mathbb{Z}$	*Integrality*

The Knapsack Problem
Symbolic Model

S	Set of product types
x_i	Number of items from each type
R_i	Revenue from each unit of type i
W_i	Weight of each unit of type i
V_i	Volume of each unit of type i
W_{MAX}	Total weight capacity of the truck
V_{MAX}	Total volume capacity of the truck
maximize	$\sum_{i \in S} R_i x_i$
subject to	$\sum_{i \in S} W_i x_i \leq W_{MAX}$ $\sum_{i \in S} V_i x_i \leq V_{MAX}$
	$x_i \geq 0 \quad \forall i \in S$ $x_i \in \mathbb{Z} \quad \forall i \in S$

How to solve mathematical models examples.

Question 1

The shaded area in the figure below represents the feasible region of a linear programming problem. The (unknown) objective function is to be maximised. Decide for each the following statements whether it is true or false. In each case, give an example objective function to justify your answer.

1. The point (0, 0) cannot be an optimal solution. false: Min(x_1+x_2)
2. If the objective function has a larger value at point (3, 3) than at points (0, 2) and (6, 3), then (3, 3) must be an optimal solution. True: Max$(-x_1+6x_2)$
3. If point (3, 3) is an optimal solution and multiple optimal solutions exist, then either (0, 2) or (6, 3) must also be an optimal solution.

(c) True.

$x_2=3$

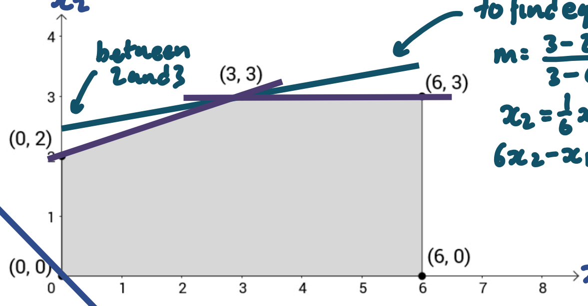
Max(x_2)

$m=\frac{3-2}{3-0}=\frac{1}{3}$

$x_2=\frac{1}{3}x_1+1$

$3x_2-x_1=3$

Max$(3x_2-x_1)$



(b) to find eqn of this line:

$m=\frac{3-2.5}{3-0}=\frac{1}{6}$

$x_2=\frac{1}{6}x_1+1$

$6x_2-x_1=6$

keep +ive! (not doing like: $\frac{1}{6}x_1-x_2=-1$) cause we be confused!

(a) to find eqn of this line:

$m=-1$

$x_2=-x_1+1$

Min(x_1+x_2) or Max$(-x_1-x_2)$

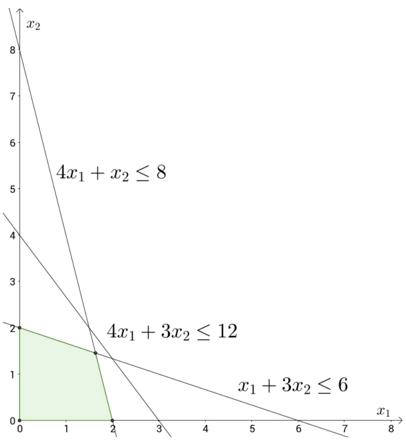
$x_2+x_1=1$ (Max) ✓ not $-x_1-x_2=-1$ (Max) X

Question 2

Using the graphical method, determine which of the following three constraints is redundant. Assume that non-negativity conditions for the decision variables x_1 and x_2 also apply.

$$x_1+3x_2\le 6 \qquad 4x_1+3x_2\le 12 \qquad 4x_1+x_2\le 8$$

Solution 2



The constraint b) $4x_1+3x_2\le 12$ is redundant, because removing it would not change the feasible region.

Question 3

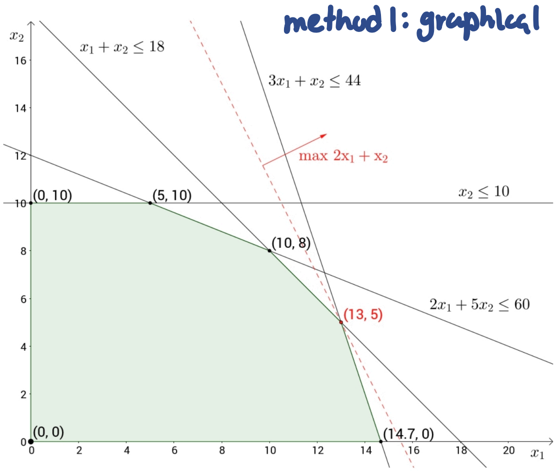
Using the graphical method solve the following problem:

maximize $2x_1 + x_2$

subject to
$$x_2 \le 10$$
$$2x_1 + 5x_2 \le 60$$
$$x_1 + x_2 \le 18$$
$$3x_1 + x_2 \le 44$$
$$x_1 \ge 0, \quad x_2 \ge 0$$

Solution 3

method 1: graphical



method 2: corner point evaluation.

Corner point	Objective value
(0, 0)	0
(0, 10)	10
(5, 10)	20
(10, 8)	28
(14.7, 0)	29.3
(13.5, 5)	31

Mathematical Modelling Examples

Question 4 (simple min cost prob. without transit)

A car manufacturer has two factories that produce buses, which are then sold to three local authorities. The production costs are the same at the both factories, while the cost of shipping each bus is shown below:

	Local Authority		
	1	2	3
Factory A	£800	£700	£400
B	£600	£800	£500

A total of 60 buses are produced and shipped per week. Each factory can produce and ship up to 50 trucks per week, so there is flexibility on how to divide the production between the two so as to reduce shipping costs. However, each local authority must receive exactly 20 buses per week. We seek to determine how many buses to produce at each factory, and how will they be shipped in order to minimize costs.

Formulate a model for this problem (the explicit form would suffice).

Solution 4.a

Decision variables: $x_{A1}, x_{A2}, x_{A3}, x_{B1}, x_{B2}, x_{B3}$

Objective function: min $800x_{A1} + 700x_{A2} + 400x_{A3} + 600x_{B1} + 800x_{B2} + 500x_{B3}$

Constraints:
$$x_{A1} + x_{A2} + x_{A3} \le 50$$
$$x_{B1} + x_{B2} + x_{B3} \le 50$$
$$x_{A1} + x_{B1} = 20$$
$$x_{A2} + x_{B2} = 20$$
$$x_{A3} + x_{B3} = 20$$
$$x_{A1}, x_{A2}, x_{A3}, x_{B1}, x_{B2}, x_{B3} \ge 0$$

transportation prob.

Question 5 (min cost prob. without transit)

An airline is expanding and would like to offer direct flights from London to Dublin, Zurich, and Frankfurt. The estimated fares and expected number of passengers and the characteristics of the two types of aircraft available are provided in the tables below.

The airline's objective is to maximise profits by selecting the optimum number of aircrafts from each type that must be deployed, and the number of return flights that must be scheduled for each destination. The airline is very keen to offer at least as many seats as required to meet current levels of demand.

a. Formulate an explicit model formulation for this problem.
b. Find a solution to this problem using Excel Solver.

Destination	Fare (one way)	Demand outbound	Demand inbound
Dublin	£175	650 p / day	550 p / day
Zurich	£230	450 p / day	500 p / day
Frankfurt	£200	760 p / day	700 p / day

Aircraft type	Seating capacity	Max utilisation	Capital Expenditure	Return flight to Dublin		Return flight to Zurich		Return flight to Frankfurt	
				Duration	OpEx	Duration	OpEx	Duration	OpEx
A320	150	12 h / day	£23,000 / day	3.5 h	£31,000	5.0 h	£47,000	4.5 h	£42,000
A321	230	11 h / day	£32,000 / day	2.5 h	£33,000	4.0 h	£59,000	3.0 h	£51,000

Objective function:

The objective of this problem is to maximise the profit. Where: $Profit = Revenue - Cost$

Destination	Revenue=Fare *demand	Cost	
		Operational Expenditure	Capital Ex. (independent of destination)
Dublin	$175 * (650 + 550)$	$31000 * x_{320}^{DUB} + 33000 * x_{321}^{DUB}$	$23000 * n_{320} + 32000 * n_{321}$
Zurich	$230 * (450 + 500)$	$47000 * x_{320}^{ZUR} + 59000 * x_{321}^{ZUR}$	
Frankfurt	$200 * (760 + 700)$	$42000 * x_{320}^{FRA} + 51000 * x_{321}^{FRA}$	

Explicit model formulation for this problem:

Decision variables: n_{320}, n_{321} (number of aircraft from each class)

$x_{320}^{DUB}, x_{320}^{ZUR}, x_{320}^{FRA}$ (flights with specific aircraft per destination)

$x_{321}^{DUB}, x_{321}^{ZUR}, x_{321}^{FRA}$

Objective function:

$$maximize \ (175 * (650 + 550) + 230 * (450 + 500) + 200 * (760 + 700)) - (23000 * n_{320} + 32000 * n_{321}) - (31000 * x_{320}^{DUB} + 33000 * x_{321}^{DUB}) - (47000 * x_{320}^{ZUR} + 59000 * x_{321}^{ZUR}) - (42000 * x_{320}^{FRA} + 51000 * x_{321}^{FRA})$$

Constraints:

$$subject\ to \quad \left.\begin{aligned} 150 * x_{320}^{DUB} + 230 * x_{321}^{DUB} &\geq 650 \\ 150 * x_{320}^{ZUR} + 230 * x_{321}^{ZUR} &\geq 500 \\ 150 * x_{320}^{FRA} + 230 * x_{321}^{FRA} &\geq 760 \end{aligned}\right\} \text{capacity: passenger demand (usually} \geq)$$

$$\left.\begin{aligned} 3.5 * x_{320}^{DUB} + 5 * x_{320}^{ZUR} + 4.5 * x_{320}^{FRA} &\leq 12 * n_{320} \\ 2.5 * x_{321}^{DUB} + 4 * x_{321}^{ZUR} + 3 * x_{321}^{FRA} &\leq 11 * n_{321} \end{aligned}\right\} \text{capacity: time supply (usually} \leq)$$

$$n_{320}, n_{321}, x_{320}^{DUB}, x_{320}^{ZUR}, x_{320}^{FRA}, x_{321}^{DUB}, x_{321}^{ZUR}, x_{321}^{FRA} \geq 0 \ \} \text{ non-negativity.}$$

Question 2

A manufacturer has two factories at A and B and seeks to transport goods to their retailers in F, G and H. The following network can be used to represent their supply chain.

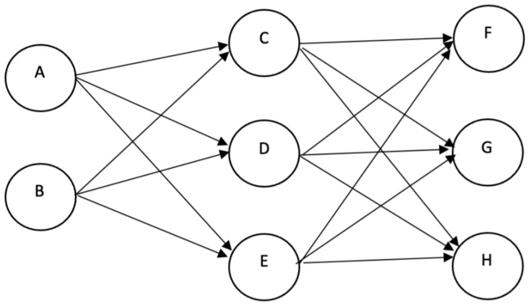
transhipment prob

(min cost prob. with transit)

a) Formulate the problem that minimises the cost of transporting cargo G1 from the supply points to the retailers, ensuring that all demand is satisfied.
b) Given the cost table below, solve the above problem using PuLP.

(min cost multicommodity prob. with transit)

c) The manufacturer seeks to extend their product line-up and expects to introduce two additional products into the supply chain (G2 and G3). Re-formulate the problem to incorporate the new commodity types, ensuring link capacities are not exceeded.



Link	Cost	Capacity
AC	5	150
AD	8	150
AE	7	100
BC	9	100
BD	5	150
BE	4	100
CF	3	75
CG	4	90
CH	8	150
DF	7	130
DG	5	100
DH	6	120
EF	9	160
EG	4	90
EH	4	90

Node	Demand		
	G1	G2	G3
A	75	25	60
B	75	45	40
C	0	0	0
D	0	0	0
E	0	0	0
F	-50	-30	-50
G	-60	-15	-40
H	-40	-25	-10

Solution 2

a.

Sets	N	Set of nodes
	E	Set of edges
Parameters	c_j	Transportation costs of link (i,j)
	D_i	Supply/Demand of each node i
	L_{ij}	Link capacity
Decision Variables	x_{ij}	Integer: Number of goods transported at each link (i,j)

$$minimise \sum_{(i,j)\in E} x_{ij}c_{ij}$$

Subject to

$$\sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = D_i \quad \forall i \in N$$

capacity constraints of each arc →

$$x_{ij} \le L_{ij} \quad \forall (i,j) \in E$$

$$x_{ij} \ge 0 \quad \forall (i,j) \in E$$

b. Solution in notebook provided.

c.

$$minimise \sum_{(i,j)\in E} x_{ij}^l c_{ij}$$

Subject to

$$\sum_{j:(i,j)\in E} x_{ij}^l - \sum_{j:(j,i)\in E} x_{ji}^l = D_i \quad \forall i \in N, l \in C$$

$$\sum_{l\in C} x_{ij}^l \le L_{ij} \quad \forall (i,j) \in E$$

$$x_{ij}^l \ge 0 \quad \forall (i,j) \in E, l \in C$$

C4 Introduction to Logistics

Introduction to Logistic

Definitions.

- Logistics vs Supply Chain Management:

Logistic focus on efficient MOVEMENT and STORAGE of goods.

SCM is a superset of supply, production and distribution

- Potential issues in Logistics: {C,T,D}
  - Excess costs
  - Poor timing (need more storage)
  - Upstream disruptions (stagnation supply)
  - Downstream disruptions. (affect information flow)

- Mode of transport:

Air: High value, urgent, perishable goods

Sea/Rail: Bulk, heavy, non-perishable goods

- Why hold inventory (from inventory management)
  - lead time (buffer against delay)
  - uncertainty (demand fluctuation)
  - reliability
  - economic of scale (bulk buying)
  - speculation.

{R,U,L,E,S}

Inventory Management*

I am orderer, how much to order?

1. Economic Order Quantity (EOQ) (used when stock arrives instantly)

- to find optimal order volume, Q^* and reorder frequency, D/Q^* that balances ordering and holding cost.

Total Cost = Ordering Cost + Holding Cost

$$= \frac{DS}{Q} + \frac{Qh}{2}$$

$$\frac{dT}{dQ} = 0$$

D: Demand Rate

S: Order/Setup Cost per order

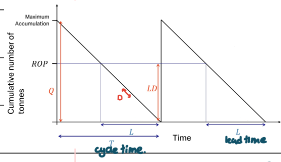
Q: Order volume

h: Holding cost per unit

($h = I \times C$, I is handling %, C is unit cost, sometimes given h straightaway)

(EOQ) Optimal Order Volume, $Q^* = \sqrt{\frac{2SD}{h}}$

- reorder freq $= D/Q^*$ (order per year)
- cycle time $= Q^*/D$ (years)
- reorder point, ROP $= L \times D$ L: lead time



I am producer, how much to produce?

2. Economic Production Quantity (EPQ) (used when stock replenish gradually over time)

- to find optimal production volume, Q^* that balances producing and holding costs.

Total Cost = Producing Cost + Holding Cost

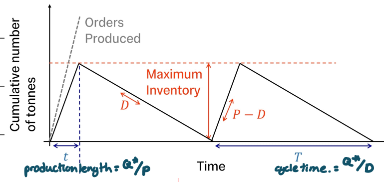
$$\frac{dT}{dQ} = 0 \qquad = \frac{DS}{Q} + \frac{Qh}{2}(1-x), \quad x = \frac{D}{P}$$

P: production rate.

(EPQ) Optimal Production Volume, $Q^* = \sqrt{\frac{2SD}{h(1-x)}}$

- Maximum inventory $= Q^* \times (1-x)$
- Production length $= Q^*/P$
- Cycle time $= Q^*/D$
- number of production $= D/Q^*$

> h/2 is average holding cost, hence eg:
>
> $$T(\text{production}) = \underbrace{\frac{D}{Q}\cdot S}_{\substack{\text{setup} \\ \text{production vol.} \times \text{setup costs}}} + \underbrace{\frac{h}{2}Q(1-x)}_{\substack{\text{holding} \\ \text{average holding cost} \times \text{inventory}}}$$



Logistic Costs

Cost of Movement

1. Handling: packaging, loading and unloading
2. Transportation: cost = $C_f + C_v q$
   - (C_f fixed cost: fleet composition, staff schedule)
   - (C_v var. cost: driver wages*, fuel and depreciation)

Cost of Holding.

1. Rent (cost maintaining the facility)
   - Service cost (stock management and insurance)
   - storage cost* (space, handling warehouse)
2. Wait (Losses incurred by delays pushing the product to the market)
   - capital cost (opportunity cost: money tied to unsold inventory)
   - risk cost (consequent of theft, deterioration, damage)

EOQ's Examples

Question 2

demand rate, D
=120000unit/year.

A company wants to determine their inventory management policy. Their annual consumption is 120,000 units, with order costs rising to £500, unit costs are £50 and handling cost is 10% of unit costs.

order cost per unit, S   unit cost, C   handling %, I

a. The optimal order volume.
b. Number of orders required per year.
c. Total ordering, total holding costs, and inventory.

(a) Handling (I) is expressed as a percentage of unit costs, and therefore $I = 0.1$. The unit cost is 50, and therefore the holding costs are: $h = IC = 50 \times 0.1 = 5$

Optimal Order Volume: $Q^* = \sqrt{2SD/h} = \sqrt{2 \times 120000 \times 500/5} = 4899$

(b) Orders per year: $D/Q^* = 120000/4899 = 24.5$ per years

Cycle Time : $Q^*/D = 4899/\ 120000 \times 365 = 14.9$ days

(c) Total ordering costs: $O = Orders\ per\ year \times S = 500 * 24.5 = 12247$

Total holding costs: $H = hQ/2 = 5 \times 4899/2 = 12247$

Inventory costs: $O + H = 12247 + 12247 = 9798$

$T = O + H = \frac{DS}{Q} + \frac{Qh}{2}$, $Q = Q^*$

(this question need to be careful with units)

Question 3

The demand for product A is 800 units every 2 weeks. The setup cost for placing an order to replenish inventory is £50. Depreciation of the product is accounted by an annual holding cost of £5 per unit. Transportation costs rise to £0.20 per unit with 5-day lead-time, increasing unit costs to £1.5/unit.

S   D   time unit for D   L

* IMPORTANT this is not C i.e. unit cost, but this is already holding cost, h (but unit is per year)

a. Determine the economic order quantity.
b. Calculate order cycle.
c. Calculate the reorder level.

Solution

since D is in bi-weekly unit, convert h from year unit to bi-weekly unit.

(a) Bi-weekly Holding costs: $h = 5 \times 2/52 = 0.192$

Optimal Order Volume: $Q^* = \sqrt{2SD/h} = \sqrt{2 \times 50 \times 800/0.196} = 645$ ← no time unit.

(b) Order Cycle: $Q^*/D = 645/800 = 0.8\ bi-weeks = 11.3\ days$

(c) Reorder Point: $ROP = \underbrace{800/14}_{unit\ day^{-1}} \times \underbrace{5}_{day} = \underbrace{286}_{unit}$ – remember to change demand to daily demand.

EPQ's Examples

Question 5

A company produces 100,000 units yearly and estimate that their annual demand is 85,000 units. Production setup costs rise to £500, with holding costs approximating £50. This company operates 270 days a year.

a. The optimal production volume.
b. Number of production cycles required every year and their length.
c. Total ordering, holding costs, and inventory costs.

Solution

(a) Optimal Order Volume: $Q^* = \sqrt{2SD/h(1-\frac{D}{P})} = \sqrt{2 \times 85000 \times 500/50(1-\frac{85000}{100000})} = 3367$

(b) Production cycles per year: $\frac{D}{Q^*} = \frac{85000}{3367} = \sim 25$

Production Length: $\frac{Q^*}{P} = \frac{3367}{100000} \times 270 = 9.1\ days$

(c) Total production costs: $O = S \times \frac{D}{Q^*} = 500 \times \frac{85000}{3367} = 12624$

Total holding costs: $H = hQ(1-\frac{D}{P})/2 = 50 \times 3367(1-\frac{85000}{100000})/2 = 12624$

Inventory costs: $O + H = 12624 + 12624 = 25248$

$T = O + H = \frac{DS}{Q} + \frac{Qh}{2}(1-x)$, $Q = Q^*$

c5 Freight Transport Network

Freight Transport Network
(Move goods efficiently)

Time Series Forecasting.
(Estimate demand, D that we used in c4)

1. Definition
- Level (a): Baseline where D oscillate.
- Trend (b): Rate of change.
- Seasonality (S): Periodic Movement tied to calender.
- Cyclical Movement (c): Periodic cycle event not tied to calender
- Random Fluctuation (e): Noise

2. Forecasting Methods. *

1. Moving average: Average of last M obs. *

$$\hat{y}_{t+1} = \frac{\sum \text{last } M \text{ observation}}{M}$$

(smooths out noise but lags behind trends)

2. Simple Exponential Smoothing: Used when no trend and no seasonality only level, a *

$$\hat{y}_{t+1} = a y_t + (1-a)\hat{y}_t$$

(high a makes it responsive but unstable.
low a make it smooth but lags behind)

3. Holt's Model: Adds a trend component, a, b
4. Winter's Model: Adds a seasonality component, a, S (but no trend)
5. Holt-Winter's Model: Adds both trend and Seasonality on top of level, a, b, S

3. Forecasting Error (pp2022q2b)
1. Error: E_t = Forecast - Actual ($F_t - D_t$)
2. Mean Squared Error: $\frac{1}{n}\sum E_t^2$ (MSE)
(penalised large error heavily)
3. Mean Absolute Percentage Error: $\frac{100}{n}\sum\left|\frac{E_t}{D_t}\right|$ (MAPE)
4. Mean Absolute Deviation: $\frac{1}{n}\sum|E_t|$ (MAD)
(Magnitude of error)
5. Bias: $\sum E_t$
(sum of error)
6. Tracking Signal: $\frac{\text{bias}}{\text{MAD}}$

Facility Location (Placing the Warehouse)

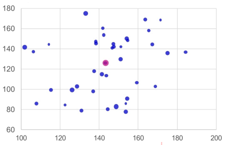
min distance

1. continuous problem (center of gravity)
→ one facility only.

Objective: Minimise the sum of weighted distances between customer locations and the coordinates of the facility (as determined by the problem).

D	*Set of customers*
x, y	*Coordinate values of facility*
w_i	*Weight of customer i*
x_i, y_i	*Coordinate of customer i*

minimize $\sum_{i\in D} w_i\sqrt{(x-x_i)^2+(y-y_i)^2}$



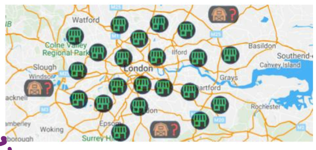
$\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$, $i \in D$

2. Single-Facility Network Location Problem
→ one facility only, and each customer served by one, and infinite capacity

Single-Facility Network Location Problem

In its simplest form, the network facility location problem needs to pick a single depot, among a set of candidates.

D	*Set of customers*
S	*Set of candidate facility locations*
x_{ij}	*Is customer i served by facility j?*
y_j	*Is a facility open in location j?*
d_{ij}	*Distance from customer i and facility j*



minimize $\sum_{i\in D}\sum_{j\in S} x_{ij}d_{ij}$

if replace d_{ij} with c_{ij} it becomes min cost prob.

subject to

$\sum_{j\in S} x_{ij} = 1 \quad \forall i \in D$ ← Every customer i will be served by a single facility j

$\sum_{j\in S} y_j = 1$ ← We require one and only one facility j to be open

$\sum_{j\in S} y_j := n$

$x_{ij} \le y_j \quad \forall i \in D, j \in S$ ← Facility j is only allowed to serve a customer if it is open

$x_{ij}, y_j \in \{0,1\} \quad \forall i \in D, j \in S$ ← x_{ij} and y_j are boolean variables

3. Multi-Facility Network Location Problem
→ multiple facility, but each customer served by one and infinite capacity.

min cost

4. Capacitated Facility Location Problem
→ flexible no. of facility, each customer can be served by multiple facility, but finite capacity.

Capacitated Facility Location Model – Mathematical Model

D	Set of customers
S	Set of candidate facility locations
x_{ij}	Quantity supplied to i served by j
y_j	Is a facility open in location j?
t_{ij}	Unit shipping cost from j to i
n_{min}	Minimum number of open locations
n_{max}	Maximum number of open locations
c_j	Facility j operating cost (if open)
Q_i	Material demand by customer i
P_j	Total supply capacity in facility j

minimize $\sum_{i\in D}\sum_{j\in S}\overbrace{x_{ij}t_{ij}}^{\text{transport cost}} + \sum_{j\in S}\overbrace{c_j y_j}^{\text{fixed opening cost}}$ (1)

subject to

$\sum_{j\in S} x_{ij} = Q_i \quad \forall i \in D$ (2)

demand satisfaction: sum of good arriving at customer i must equal to their demand (sum of supply provided meet demand)

capacity & logic:

$\sum_{i\in D} x_{ij} \le P_j y_j \quad \forall j \in S$ (3)

(sum of demand required don't exceed supply available)

logic: if facility j is closed ($y_j=0$), RHS = 0 → all flow (x_{ij}) must = 0

capacity: if facility j is open ($y_j=1$), total flow out cannot exceed supply capacity (P_j)

$n_{min} \le \sum_{j\in S} y_j \le n_{max}$ (4)

$x_{ij} \ge 0 \quad \forall i \in D, j \in S$ (5)

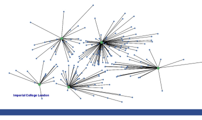
$y_j \in \{0,1\} \quad \forall i \in D, j \in S$ (6)

- The objective function (1) seeks to minimize the total cost of facility operation and material distribution
- x_{ij} is not boolean, but rather defines the quantity of material flowing from facility j to customer i
- The sum of materials transported to any customer i must be equal to their demand Q_i (2)
- The sum of materials transported from any facility j shall not exceed their total capacity P_j (3)
- If a facility j is not open ($y_j = 0$), then all flows x_{ij} for j shall be equal to 0. (also 3)
- Constraint (4) limits the upper and lower number of open facilities – will be implemented as 2 separate constraints

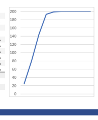
Level of Service (How much facility we act. need?)

Level of Service Calculations – 5 Facilities

Facilities	Σ Distance	Min. Distance	Max. Distance	Avg. Distance
1	7739	1	96.7	38.7
2	5489	0	78.4	27.4
3	5000	0	58.4	25.0
4	4640	0	56.1	23.2
5	4521	0	56.1	22.6



Range	Count	%
0 - 10	25	13%
0 - 20	80	40%
0 - 30	144	72%
0 - 40	193	97%
0 - 50	199	100%
0 - 60	200	100%
0 - 70	200	100%
0 - 80	200	100%
0 - 90	200	100%
0 - 100	200	100%
0 - 25	115	58%
25 - 50	84	42%
50 - 75	1	1%
75 - 100	0	0%



Time Series Forecasting Examples

Question 1

The demand for a product in each of the last five months is shown below.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Demand	5714	6446	7511	7949	8660	9335	9256	10113	11210	11568	11586	

moving average

a. Use a four-month moving average to generate a forecast for demand for every month. What is your forecast for month 12?

simple exponential smoothing

b. Apply simple exponential smoothing using a smoothing constant of 0.9 to generate a monthly demand forecast.

c. Apply simple exponential smoothing using a smoothing constant of 0.8 to generate a monthly demand forecast.

forecasting error

d. What forecasting method better represents the demand pattern shown? Discuss.

Solution 1

a.

Start forecasting at Month 5, as we are using a 4-month average. We can use an equal weighted forecast:

$$\hat{y}_t = \frac{(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4})}{4}$$

eg. $\hat{y}_5 = \frac{y_4+y_3+y_2+y_1}{4} = \frac{5714+6446+7511+7949}{4}$

$= 6905$

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Demand					6905	7642	8364	8800	9341	9979	10537	11119

Students are encouraged to test different weighting factors.

b.

since no $\hat{y}_1$, $\hat{y}_1 = y_1$

Start forecasting at Month 2 using the following equation:

$$\hat{y}_t = 0.9y_{t-1} + 0.1\hat{y}_{t-1}$$

eg. $\hat{y}_2 = 0.9y_1 + 0.1\hat{y}_1$
$= y_1$
$= 5714$

$\hat{y}_3 = 0.9y_2 + 0.1\hat{y}_2$
$= 0.9(6446) + 0.1(5714)$
$= 6372.8$

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Demand		5714	6373	7397	7894	8583	9260	9256	10028	11092	11520	11579

c.

Start forecasting at Month 2 using the following equation:

$$\hat{y}_t = 0.8y_{t-1} + 0.2\hat{y}_{t-1}$$

eg. $\hat{y}_3 = 0.8y_2 + 0.2\hat{y}_2$
$= 0.8(6446) + 0.2(5714)$
$= 6300$

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Demand		5714	6300	7269	7813	8490	9166	9238	9938	10955	11445	11558

d.

For the period between April and November, the performance indicators yield:

Method	MSE	MAPE	MAD	BIAS	TS
Moving Averages	2226704	14.3	1452	-10161	-7
Exponential (0.9)	565105	7.4	653	-6517	-10
Exponential (0.8)	668244	8.2	730	-7304	-10

(not really recommend to revise due to complication)

Facility Location Problem Examples

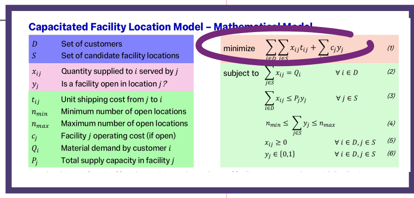
instead, reccommended to revise 2023 q2c

Question 1 (this example is slightly too complicated cause of the parameter z !)

	AA	BB	CC	DD	Yearly trips
A	£150	£250	£200	£25	40
B	£150	£250	£200	£75	35
C	£75	£200	£150	£125	100
D	£150	£200	£125	£125	25
E	£100	£200	£125	£150	40
F	£175	£175	£125	£125	25
G	£150	£175	£100	£150	50
H	£150	£150	£100	£200	30
Fixed Cost	£43935	£54892	£57820	£31471	-

A consulting company is seeking to establish its new headquarters. They have selected four possible locations AA, BB, CC, and DD, which would each incur a fixed cost described in the above table. Their clients are in 8 regions (A-H).

a) Select the optimal location for a single facility that minimises the total cost.
b) Formulate the multiple-facility location problem that minimises the total cost, considering that each consultant can take 10 trips at most annually.
c) Solve the above problem using PuLP, using the accompanying notebooks provided.
d) How would you modify your above formulation if:
   i) Only 20 consultants can be assigned to each office.
   ii) Each customer can only be serviced by one office.
   iii) Both of the above conditions apply.
e) Solve the mathematical problem described in the above sub-question. Does the answer change substantially from sub-question c)?



a. Calculate the total cost for facility location using the following equation:

$$Fixed\ Costs + \sum_{i\in F}(yearly\ trips) \times (travel\ cost) = Total\ Yearly\ Cost$$

The cost of each facility is therefore:

	AA	BB	CC	DD
Total Cost	86810	124267	107070	73346

DD is the optimal location.

b. The formulation is:

Sets	S	Set of suppliers
	D	Set of customers
Parameters	c_j	Fixed cost of establishing an office at site j
	t_{ij}	Travel cost between each site j and customer i
	Q_i	Demand of each customer i
	P	Consultant trip capacity
	M	Big number
Decision Variables	x_{ij}	Integer: Number of trips allocated from each site j to each client i
	y_j	Boolean: Whether office is established at each site j
	z_j	Integer: Number of consultants hired at each office

$$minimise \sum_{i\in D}\sum_{j\in S} x_{ij}t_{ij} + \sum_{j\in S} c_j y_j$$

Subject to

$$\sum_{j\in S} x_{ij} = Q_i \quad \forall i \in D$$

$$\sum_{j\in S} y_j \ge 1$$

$$y_j M \ge z_j \quad \forall j \in S$$

$$\sum_{i\in D} x_{ij} \le P z_j \quad \forall j \in S$$

$$x_{ij} \ge 0 \quad \forall i \in D, \forall j \in S$$

$$y_j \in \{0,1\} \quad \forall j \in S$$

$$z_j \ge 0 \quad \forall j \in S$$

d. i) $z_j \leq 20 \quad \forall j \in S$

ii) Create a new variable that indicates if office is assigned to demand $v_{ij} \in \{0,1\}$

$$\sum_{j \in S} v_{ij} = 1 \quad \forall i \in D$$

$$x_{ij} \leq v_{ij} M \quad \forall i \in D, \forall j \in S$$

Question 2 (also too complicated for revision, due to type f parameter)

A manufacturer intends to open a business in a new country. Over the next year, they expect to sell 250,000 units in the North, 140,000 in the West, 180,000 in the South, and 110,000 in the East. Three types of factories can be built: Small, Medium, and Large, which can produce 200,000, 300,000 and 400,000 units of material every year. These can be built at four locations: A, B, C, and D.

a) Formulate the facility location problem that minimises the cost of satisfying the demand.
b) Given the table of costings below, solve the above problem using PuLP.

	A	B	C	D
Annual Fix Cost Small (£ million)	6	5.5	5.6	6.1
Annual Fix Cost Medium (£ million)	8.5	7.8	7.9	8.7
Annual Fix Cost Large (£ million)	10	9.2	9.3	10.2
North	£211	£232	£238	£299
West	£232	£212	£230	£280
South	£240	£230	£215	£270
East	£300	£280	£270	£225

Solution 2

a.

Sets	S	Set of suppliers
	D	Set of customers
	F	Set of factory types
Parameters	c_{jf}	Fixed cost of establishing a factory of type t at site j
	t_{ij}	Transportation cost between each site j and customer i
	Q_i	Demand of each customer i
	P_f	Production capacity of factory type f
Decision Variables	x_{ij}	Integer: Number of goods allocated from each site j to each client i
	y_{jf}	Integer: Number of factories of type t are established at each site j

$$minimise \sum_{i \in D} \sum_{j \in S} x_{ij} t_{ij} + \sum_{j \in S} \sum_{f \in F} c_{jf} y_{jf}$$

Subject to

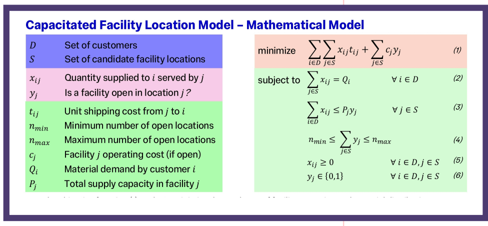
$$\sum_{j \in S} x_{ij} = Q_i \quad \forall i \in D$$

$$\sum_{j \in S} \sum_{f \in F} y_{jf} \geq 1$$

$$\sum_{i \in D} x_{ij} \leq \sum_{f \in F} P_f y_{jf} \quad \forall j \in S$$

$$x_{ij} \geq 0 \quad \forall i \in D, \forall j \in S$$

$$y_{jf} \geq 0 \quad \forall j \in S, \forall f \in F$$



C6. Touring and Routing

VRP Family

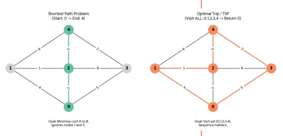
CVRP	DVRP	VRPTW
Constrained Vehicle Routing Problem	Distance - Constrained Vehicle Routing Problem	Vehicle Routing Problem with Time Windows
Vehicles have limited cargo capacity	*Vehicles have a specified range, which must not be violated*	*Vehicles must visit customers during a specific time slot*

MDVRP	PVRP	SDVRP	VRPB
Multiple Depot VRP	Periodic VRP	Split Delivery VRP	VRP with Backhauls
Same as VRP, but customers are also assigned to depots	*The classical VRP deals with a single day. The PVRP extends planning over multiple days*	*The VRP serves each customer with a single vehicle. SDVRP involves multiple vehicles per customer, if it is cheaper.*	*After delivering cargo to all customers, vehicles visit others to pick up goods for return to depot*
VRPPDP	**IRP**	**GVRP**	**HVRP**
VRP with Pickup and Delivery	Inventory Routing Problem	Green Vehicle Routing Problem	Heterogenous Fleet Vehicle Routing Problem
Same as VRPB, but backhauls (pick-ups) can take place while making deliveries	*Manages deliveries to customers over a longer horizon, while monitoring their inventories and dispatching cargo, accordingly*	*Plans deliveries in a way that minimises energy consumption*	*A non-uniform fleet is used, involving vehicles of various capacities.*

memorise a few ~4 of these VRPs
CVRP (= capacitated)
- DVRP (distance)
- VRPTW (time window)
- MDVRP (multiple depot)
- IRP (inventory)
- GVRP (green - minimise energy)

Touring and Routing

Touring (Single vehicle, visit all nodes)



touring ≠ shortest path prob.
touring visit all nodes
and shortest path doesn't
but both tries to save costs.

no. of tour = $(n-1)!/2$

1. Travelling Salesman Problem (TSP) (one vehicle, infinite capacity)

N	Set of vertices (cities)
x_{ij}	Are vertices i and j connected?
c_{ij}	Cost of travel from i to j

minimize $\sum_{i\in N}\sum_{j\in N; i\neq j} x_{ij}c_{ij}$ — minimise total cost

subject to
$\sum_{j\in N; i\neq j} x_{ij} = 1 \quad \forall i \in N$
$\sum_{i\in N; i\neq j} x_{ij} = 1 \quad \forall j \in N$
$x_{ij} \in \{0,1\} \quad \forall i,j \in N$

must leave every node once

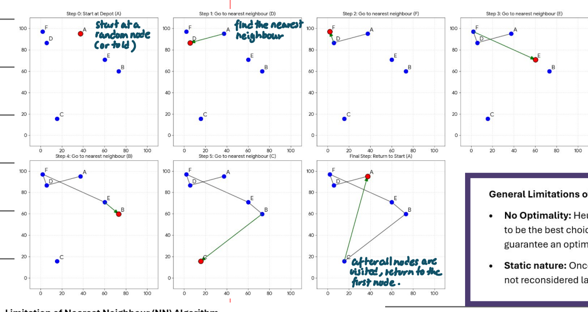
all these are subtour
mTSP, CVRP can have subtour too!
to only have one route, we have to get rid of subtour by adding constraints:

method 1: lazy constraints:
$x_{ij} + x_{ji} \le 1 \quad \forall i,j \in N$
n=1 get rid of two nodes subtour
n=2 get rid of three nodes subtour...

method 2: MTZ constraints:
$$u_i - u_j \le U(1-x_{ij}) - 1 \quad \forall \begin{cases} i,j \in N \\ i \neq j \\ i \neq 0 \\ j \neq 0 \end{cases}$$
$x_{ij} \in \{0,1\} \quad \forall i,j \in N$
$u_i \le U-1 \quad \forall i \in N$

heuristic

2. Nearest Neighbour Algorithm (Greedy Heuristic Approach, i.e. do whatever seems to be the best choice)



General Limitations of Greedy Heuristics (can be used on any heuristic qtn)

- **No Optimality:** Heuristics generally operate on the principle of "do whatever seems to be the best choice at any time and never look back." Therefore, they do not guarantee an optimal solution (unlike exact mathematical programming methods).
- **Static nature:** Once a decision is made (e.g., adding a node to a route), it is usually not reconsidered later in the process.

Limitation of Nearest Neighbour (NN) Algorithm

- g **Optimality:** There are no guarantees that an NN-tour is optimal.
- **Geometry/Crossings:** An NN-derived tour might not be convex (in Euclidean space), meaning the route paths may cross over each other, which is generally inefficient.
- **Initialization Sensitivity:** The resulting tour structure depends entirely on the choice of the starting node. A different starting node usually yields a completely different tour.
- **Greedy Nature:** It makes the locally optimal choice at each step without considering the global picture, often leading to very long legs at the end of the tour to return to the start.

• g stands for general

Routing (Multiple vehicle)

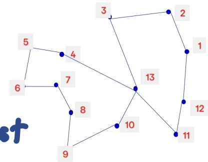
1. Multiple Travelling Salesman Problem (mTSP) (m vehicles, infinite capacity)

Multiple Travelling Salesman Problem

N	Set of vertices (cities)
x_{ij}	Are vertices i and j connected?
c_{ij}	Cost of travel from i to j
d	Index of depot vertex
m	Number of tours from d

minimize $\sum_{i\in N}\sum_{j\in N; i\neq j} x_{ij}c_{ij}$ — minimise total cost

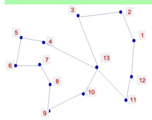
subject to
$\sum_{j\in S; d\neq j} x_{dj} = m$ — Exactly m arcs (...and tours) commence from vertex d
$\sum_{i\in S; d\neq i} x_{id} = m$ — Exactly m arcs (...and tours) terminate at vertex d
$\sum_{j\in S; i\neq j} x_{ij} = 1 \quad \forall i \in N, i \neq d$ — As in TSP, but with the exception of d
$\sum_{i\in S; i\neq j} x_{ij} = 1 \quad \forall j \in N, i \neq d$ — As in TSP, but with the exception of d
$x_{ij} \in \{0,1\} \quad \forall i,j \in N$
+ subtour elimination constraints — Subtours affect the mTSP as well – we need an appropriate set of SECs that would allow multiple tours



2. Capacitated Vehicle Routing Problem (CVRP) * (m vehicles, but finite capacity, i.e. payload p)

Capacitated Vehicle Routing Problem

V	Set of vertices
S, D	Set of suppliers, customers $S, D \subset V$
x_{ij}	Are vehicles travelling from i and j?
f_{ij}	How much cargo to carry from i to j
d_{ij}	Distance between vertices i and j
p	Vehicle payload
Q_i	Demand by customer i
c	Investment cost per vehicle
B	Budget



minimise $\sum_{i\in N}\sum_{j\in N; i\neq j} x_{ij}d_{ij}$ (1) In this example it is distance, d_{ij}. It can be cost, c_{ij} as well.

subject to $\sum_{j\in V} f_{ji} - \sum_{j\in V} f_{ij} = Q_i \quad \forall i \in D$ (2) — Cargo equilibrium: cargo in – cargo out = demand (for each node's customer)

$\sum_{j\in V} x_{ji} - \sum_{j\in V} x_{ij} = 0 \quad \forall i \in D$ (3)

* $Q_j \cancel{x_{ij}} \le f_{ij} \le (p - Q_i)\cancel{x_{ij}} \quad \forall i,j \in V$ (4) — Vehicle capacity limits tour size
$Q_j \le f_{ij} \le p - Q_i$ (next cust. demand ≤ current cargo ≤ limit - last cust. demand)

$\sum_{j\in D} x_{ij} \ge 1 \quad \forall i \in S$ (5)

$\sum_{j\in D} c\, x_{ij} \le B \quad \forall i \in S$ (6) — Bounded by a budget B

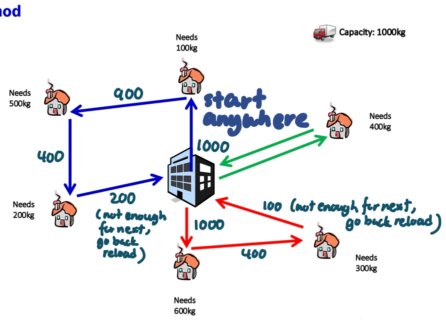
$f_{ij} \ge 0 \quad \forall i \in V, \forall j \in V$ (7) — New variable for cargo f

$x_{ij} \in \{0,1\} \quad \forall i \in V, \forall j \in V$ (8)

heuristic

3. Heuristic Method 1: The sweep method

Sweep Method
Step 11



Limitation of The Sweep Method

- **Network Applicability:** It is most applicable for **Euclidean planes**. It is **not practical for road networks** where distances are not straight lines.
- **Initialization Sensitivity:** The structure of the tours depends heavily on the choice of the **seed node** (the first customer picked). It is usually necessary to repeat the algorithm with various seeds to find the best tour.
- **Depot Location:** If the depot does not coincide with the **geometrical centroid** of the customers, the tour lengths will vary significantly. Off-centre depots can also lead to tours with crossings.
- **Capacity Waste:** Once a capacity constraint is reached, the tour is immediately terminated, and the vehicle returns to the depot. This rigidness might lead to significant waste of vehicle capacity.
- g **Optimality:** As a greedy algorithm, it is difficult to achieve optimality compared to exact mathematical approaches.
- g **Static Nature:** Once a node (customer) is added to a tour, it cannot be considered again later in the process. It locks in early decisions which might prevent better combinations later.

Limitation of The Savings Method

- g **Static Nature:** Once a node (customer) is added to a tour, it cannot be considered again later in the process. It locks in early decisions which might prevent better combinations later.
- **Spatial Sensitivity:** The method is sensitive to the spatial distribution of the hub and nodes. If the hub/depot is **not located centrally**, the resulting tours will be highly imbalanced.
- g **Optimality:** As a greedy algorithm, it is difficult to achieve optimality compared to exact mathematical approaches.

4. Heuristic Method 2: The savings method. *

The Savings Method

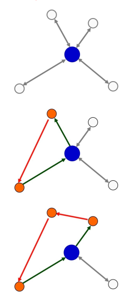
- Developed by Clarke and Wright (1963) – also an iterative approach.
- Initially considers one tour per customer (depot → customer → depot).
- Assuming that vertex 0 is the depot, the total distance of all tours is:

$$\sum_{j=1..n} (c_{0j} + c_{j0})$$

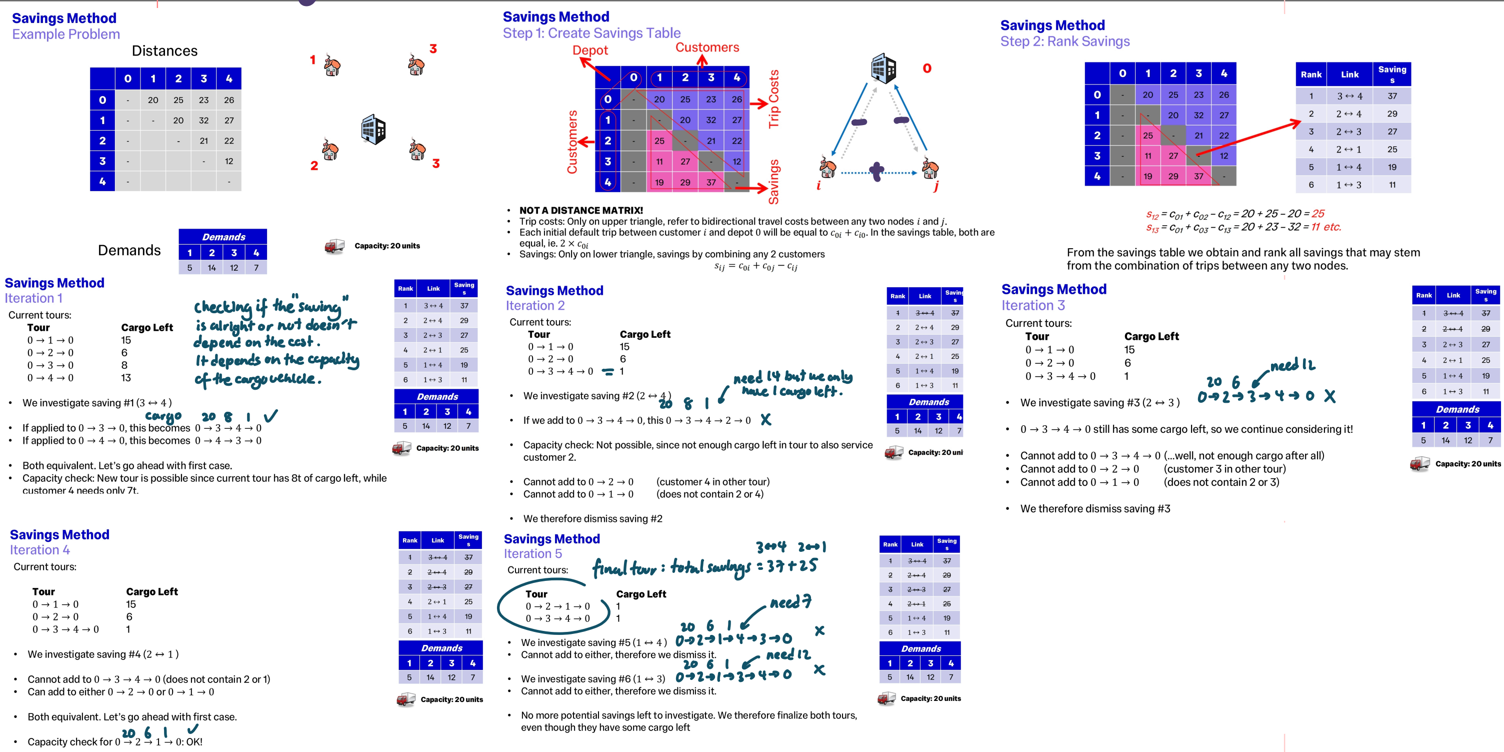
- Trip pairings are considered to reduce costs (savings):

$$s_{ij} = c_{i0} + c_{oj} - c_{ij}$$

- At each iteration we consider the largest known saving and add it to a tour whose capacity is not violated.
- Algorithm terminates when all potential savings have been evaluated.



The saving methods in action:



Touring Example: (Heuristic)

Question 3

Apply the nearest neighbour heuristic once to obtain a travelling salesman route among a set of nodes described with the distance matrix below. Comment on the optimality of the result.

		1	2	3	4	5	6	7
1.	1	0.00	72.01	9.43	91.92	88.51	76.69	106.51
3.	2	72.01	0.00	64.26	61.77	18.87	65.86	69.26
2.	3	9.43	64.26	0.00	82.64	80.13	67.90	97.17
5.	4	91.92	61.77	82.64	0.00	55.17	21.47	15.00
4.	5	88.51	18.87	80.13	55.17	0.00	65.31	58.83
7.	6	76.69	65.86	67.90	21.47	65.31	0.00	35.61
6.	7	106.51	69.26	97.17	15.00	58.83	35.61	0.00

Solution 3

Assuming that we commence our tour from Node 1, the route will be 1-> 3-> 2-> 5->4->7->6->1, with a total cost of 275.03. The Nearest Neighbour heuristic provides an approximate result with no guarantees of optimality. Furthermore, different results may be obtained depending upon the node chosen to depart from initially.

Choose initial centre points P1 = A = (2, 10) and P2 = C = (8, 4).

Routing Example : (Heuristic)

Question 2

Using the savings method, calculate a set of VRP tours that will create a product distribution plan using a fleet of vehicles, with each having a capacity of 23 units.

Assume that the depot node is 1, and the quantity of refuge units to be collected from each node is as follows:

Node	2	3	4	5	6	7	8	9	10
Quantity	4	6	5	4	7	3	5	4	4

The distances among nodes are provided as follows:

	1	2	3	4	5	6	7	8	9	10
1	-	25	43	57	43	61	29	41	48	71
2		-	29	34	43	68	49	66	72	91
3			-	52	72	96	72	81	89	114
4				-	45	71	71	95	99	108
5					-	27	36	65	65	65
6						-	40	66	62	46
7							-	31	31	43
8								-	11	46
9									-	36
10										-

Solution 2

The distance/savings matrix is as follows.

	1	2	3	4	5	6	7	8	9	10
1		25	43	57	43	61	29	41	48	71
2			29	34	43	68	49	66	72	91
3		39		52	72	96	72	81	89	114
4		48	48		45	71	71	95	99	108
5		25	14	55		27	36	65	65	65
6		18	8	47	77		40	66	62	46
7		5	0	15	36	50		31	31	43
8		0	3	3	19	36	39		11	46
9		1	2	6	26	47	46	78		36
10		5	0	20	49	86	57	66	83	

We construct a ranked savings table, as shown in the matrix below. The capacity of each vehicle is 23, and we assume that no other constraints apply.

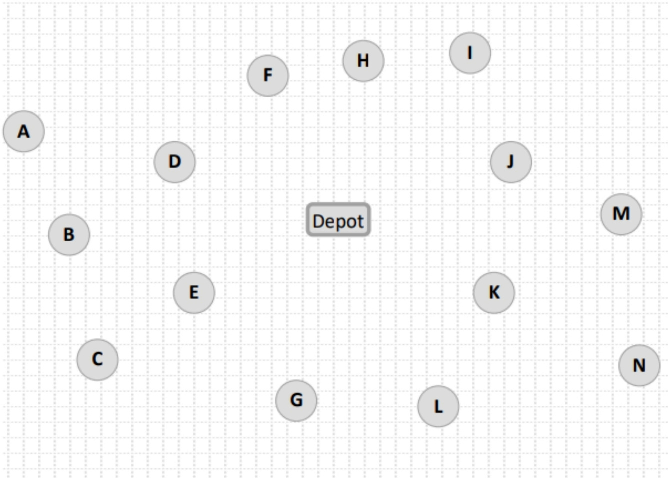
Link	Savings	Link	Savings	Link	Savings
(6, 10)	86	(6, 9)	47	(3, 5)	14
(9, 10)	83	(7, 9)	46	(3, 6)	8
(8, 9)	78	(2, 3)	39	(4, 9)	6
(5, 6)	77	(7, 8)	39	(2, 7)	5
(8, 10)	66	(5, 7)	36	(2, 10)	5
(7, 10)	57	(6, 8)	36	(3, 8)	3
(4, 5)	55	(5, 9)	26	(4, 8)	3
(6, 7)	50	(2, 5)	25	(3, 9)	2
(5, 10)	49	(4, 10)	20	(2, 9)	1
(3, 4)	48	(5, 8)	19	(2, 8)	0
(2, 4)	48	(2, 6)	18	(3, 7)	0
(4, 6)	47	(4, 7)	15	(3, 10)	0

We proceed in steps, while progressing through the ranked savings list.

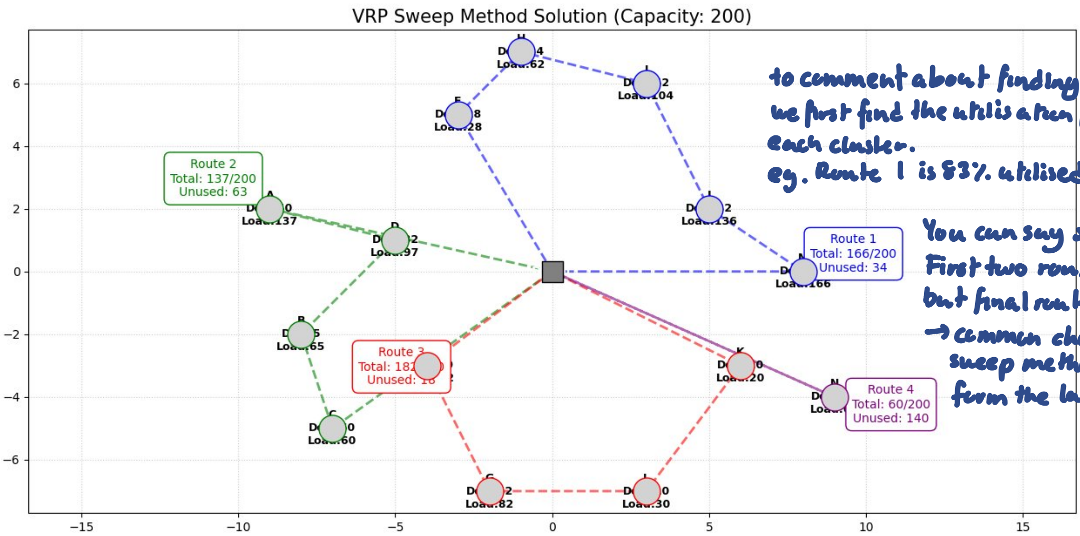
Step 1:	Add link (6, 10) to I	Current tour: I {1, 6, 10, 1}
Step 3:	Add link (9, ~~1~~10) to I	Current tour: I {1, 6, 10, 9, 1}
Step 4:	Add link (8, 9) to I	Current tour: I {1, 6, 10, 9, 8, 1}
Step 5:	Skip link (5, 6)	(it would violate capacity constraints in tour I)
Step 6:	Skip link (8, 10)	(nodes 8 and 10 already exist in tour I)
Step 7:	Skip link (7, 10)	(node 10 already exists in tour I)
Step 9:	Link (4, 5)	contains two nodes that do not exist in tour I, so we initiate tour II
	Add link (4, 5) to II	Current tour: II {1, 4, 5, 1}
Step 10:	Add link (6, 7) to I	Current tour: I {1, 7, 6, 10, 9, 8, 1}
Step 11:	Tour I now has a load of 23, and therefore we close it.	
Step 12:	Skip link (5, 10)	(nodes 5 and 10 already exist in our tours)
Step 13:	Add link (3, 4) to II	Current tour: II {1, 3, 4, 5, 1}
Step 14:	Skip link (2, 4)	(node 4 already exists in tour II)
Step 15:	Skip link (4, 6)	(nodes 4 and 6 already exist in our tours)
Step 16:	Skip link (6, 9)	(nodes 9 and 6 already exist in our tours)
Step 17:	Skip link (7, 9)	(nodes 9 and 7 already exist in our tours)
Step 18:	Add link (2, 3) to II	Current tour: II {1, 2, 3, 4, 5, 1}
Step 19:	We have now included all links to our tours, so we can now close tour II as well. Their combined length (total distance travelled) is 397 units.	

Question 3

Using the sweep method, obtain the VRP tours for the following networks. Comment on your findings and discuss the limitations of the method. Assume that each vehicle has the capacity of 200 tonnes.



Node	*Demand*
A	*40*
B	*5*
C	*60*
D	*32*
E	*100*
F	*28*
G	*52*
H	*34*
I	*42*
J	*32*
K	*20*
L	*10*
M	*30*
N	*60*



to comment about finding:
we first find the utilisation for each cluster.
eg. Route 1 is 83% utilised.

You can say solution is unbalanced.
First two route are well-utilised,
but final route is not.
→ common characteristic of sweep method (leftover nodes form the last route)

C7. Transport Planning

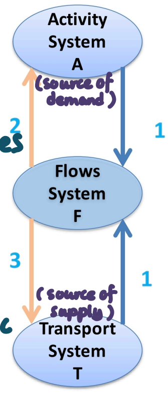
Aim: To solve the "Prediction Problem"
- How to predict the impact of an intervention?
- How to evaluate those impacts are desireable compared to the costs?

(eg: Should we widen the road? reduced time travel vs. construction cost)

Transport Planning

Manheim's Analysis Framework

- Activity (A): Landuse, socio-economic factors, where people live or work.
- Transport system (T): Physical networks, vehicles, schedules, prices.
- Flow system (F): Actual traffic, passenger volumes, travel time (resulting from A and T)



2. Long term ($F \to A$)
flow changes demand
eg. high congestion causes business to relocate

3. Long term ($F \to T$)
flow changes supply.
eg. high volume of traffic pressure government to build wider road.

long term because business can't relocate immediately, road can't be built overnight.

1. short term. ($A \to F$, $T \to F$)
A generates demand for travel, and T provides the supply/service
These interact to determine F on the network
(short-term equilibrium)

eg. Christmas, high A, but high A means low T, but low T means low A, but low A means high T, but high T means high A ...
till it reach an equilibrium.
(spiralling effect is short term!)

Mathematical Formulisation and Equilibrium.

(Manheim become mathematic's equation)

1. Short term — relationship 1 ($A \to F$, $T \to F$)

travel-demand function. $V = D(A, S)$; $S = J(V, T)$ link-performance function

- V: volume of travel demanded
- A: activity system attributes (population, income)
- S: transport system service (travel time, cost)
- T: transport system attributes (capacity, route)

note that you don't see F in the eqn because F is a vector of flow, i.e. $F = (V, S)$

2. Long term — relationship 2 and 3 ($F \to A$, $F \to T$)

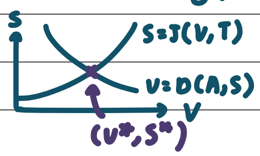
$A = H(T, F)$; $T = I(A, F)$ (not really important)

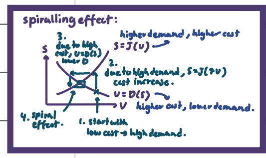
3. Demand-Performance Equilibrium *

A system is said to be equilibrium when volume of travel demanded, V matches the system's ability to supply that travel at a specific service level, S:

$V^* = D(A, S^*)$; $S^* = J(V^*, T)$ ← flow when system is in equil.

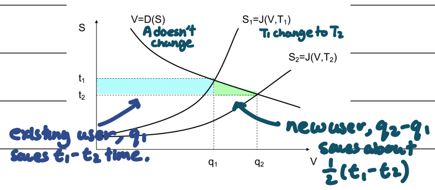
solve simultaneously for $(V^*, S^*) = F^*$:





4. Economic Evaluation and "Rule of a Half"

This answer our "Prediction Problem" i.e. if impact of an intervention are desirable compared to the cost.



total benefit (both existing and new) $= \frac{1}{2}(q_1 + q_2)(t_1 - t_2)$ (Area of trapezium)

we will check if this is greater than a threshold:

$\frac{1}{2}(q_1+q_2)(t_1-t_2) > \lambda$

Congestion and Cost

q: flow ~ volume
c: cost

Private Cost for ONE Traveller: $c(q)$

Total Social Cost for ALL Traveller: $q \cdot c(q)$

Marginal social cost (change in total social cost when you add one more traveller)

$$MSC = \frac{d}{dq}(q \cdot c(q)) \quad \text{product rule.}$$

$$= \underbrace{c(q)}_{\text{private cost}} + \underbrace{q \cdot \frac{dc}{dq}}_{\text{public cost (externality)}}$$

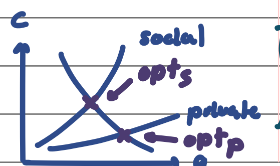
add 1 person, add 1 private cost is normal and ok

private cost (you pay your own delay)

public cost (externality) (everyone else have to pay for the slight increase in delay you caused for being there)

this term is the reason why congestion is a prob. because $\frac{dc}{dq} > 0$ (costs always rises with traffic) $\to q \cdot \frac{dc}{dq} > 0$

→ social cost is always higher than private cost.
(c at opt_p < c at opt_s)



hence for society, eg. we should take bus even if it is inconvenient (CT) cause for society its beneficial ($q \cdot \frac{dc}{dq}$)

→ another issue with c at opt_p < c at opt_s is q at opt_p > q at opt_s:
since driver doesn't pay for $q \frac{dc}{dq}$, demand high.

Demand-Performance Equilibrium Examples

INTRODUCTION

In a congestible facility, the generalised cost of travel c depends on the level of demand q. Likewise, the level of demand will depend on the generalised cost. Hence

Performance: $c = C(q)$ → $S=J(V)$

Demand: $q = Q(c)$ → $V=D(S)$

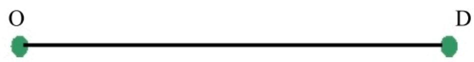
The natural approach to predicting the actual demand and cost that will be experienced by travellers on the network is to seek the simultaneous solution of these equations:

$c^* = C(q^*)$

$q^* = Q(c^*)$

The pair (q^*, c^*) is called an equilibrium flow pattern (see Manheim's paper).

In general, for large networks or complex patterns of demand, the computation of such equilibria is rather difficult. However, for simple networks, the computations can be straightforward. Consider for example, the simple single link network shown below:



Suppose that the demand for travel on this network from the origin O to the destination D is given by the following (linear) demand model:

$$q = \alpha - \beta c$$

and that the generalised cost of travel from O to D is given by the following (linear) performance model:

$$c = \mu + \lambda q$$

where the quantities $\alpha, \beta, \mu, \lambda$ are all positive parameters.

Tasks

1. Discuss the physical interpretation of the parameters .

2. Compute the equilibrium values q^* and c^*, in terms of these parameters and discuss your result.

* more congested mean $\frac{dc}{dq}$ is large!

3. How do the equilibrium values q^* and c^* change if the network becomes less congestible?

very typical demand-performance questions.

increase c-intercept

* HARD 4. Suppose that the network operator wanted to introduce a fixed toll p for travellers using this facility and wants to know what level of toll would maximise the total revenue generated by the network. How would you extend the model above to answer this question?

interesting and important to understand.

1) $V=D(S)$

The demand model is

$$S=J(V) \qquad q = \alpha - \beta c \qquad \alpha, \beta > 0 \qquad (1)$$

and the performance model is

$$c = \mu + \lambda q \qquad \mu, \lambda > 0 \qquad (2)$$

α = maximum demand possible
β = sensitivity of demand to changes in generalised cost
μ = minimum generalised cost (e.g. in uncongested conditions)
λ = sensitivity of generalised cost to changes in demand - the congestability of the network.

2)

The equilibrium values of demand and generalised cost q^* and c^* satisfies:

$$\left.\begin{aligned} q^* &= \alpha - \beta c^* \quad (3) \\ c^* &= \mu + \lambda q^* \quad (4) \end{aligned}\right\}$$

to find equilibrium solution to demand-performance model
→ solve simultaneous eqn.

$$=> \qquad q^* = \frac{\alpha - \beta\mu}{1 + \beta\lambda} \qquad c^* = \frac{\mu + \lambda\alpha}{1 + \beta\lambda} \qquad if\,(\alpha > \beta\mu) \tag{5}$$

3)

The equilibrium can be represented as the intersection of the curves representing the demand and generalised cost models

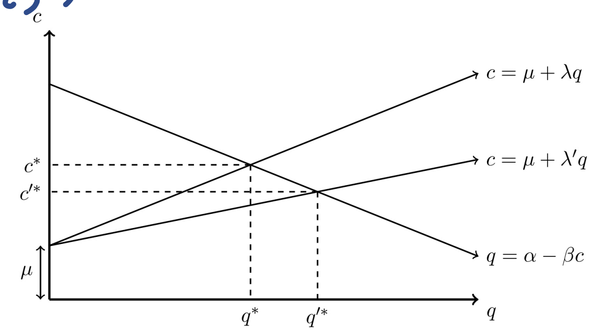
less congestable doesn't mean q reduces. It means c (travel time) doesn't increase significantly as q increases.
i.e. c is now less sensitive to q → gradient of c-q graph reduce ($\lambda \to \lambda'$)

If the network becomes less congestable, then the value of λ reduces, say to λ', and the equilibrium shift to (c'^*, q'^*) from (c^*, q^*) where $c'^* < c^*$ and $q'^* > q^*$.

Algebraically:

this basically means as λ increases (λ' to λ) q^* decreases (q'^* to q') and c^* increase (c'^* to c')

$$\left.\begin{aligned} \frac{\partial q^*}{\partial \lambda} &= \frac{-\beta(\alpha - \beta\mu)}{(1+\beta\lambda)^2} \qquad < 0 \quad (6) \\ \frac{\partial c^*}{\partial \lambda} &= \frac{\alpha - \beta\mu}{(1+\beta\lambda)^2} \qquad > 0 \quad (7) \end{aligned}\right\}$$



4) (quite hard, especially the concept of

1. toll relates directly to S (system service level)
2. Since c is about generalise cost in time units, we need to convert p's unit.

Suppose that a toll of p is charged. This has the effect of increasing the minimum generalised cost of travel from μ to $(\mu + \theta p)$, where θ is a parameter that converts units of toll (money) into units of travel time (assuming that we are measuring generalised cost in time units). θ is a parameter that characterises travel demand.

Therefore, the equilibrium with the toll p is:

$$q^* = \frac{\alpha - \beta(\mu + \theta p)}{1 + \beta\lambda} \qquad c^* = \frac{(\mu + \theta p) + \lambda\alpha}{1 + \beta\lambda} \tag{8}$$

The revenue (in units of money) that the toll raises, is $q^* \times p$

$$R = q^* \times p = \frac{p\,(\alpha - \beta\,(\mu + \theta p))}{1 + \beta\lambda} \tag{9}$$

The toll revenue is maximised when $\dfrac{\partial R}{\partial p} = 0$ and $\dfrac{\partial^2 R}{\partial p^2} < 0$

$$\frac{\partial R}{\partial p} = \frac{\alpha - \beta\mu - 2p\beta\theta}{1 + \beta\lambda} \qquad \frac{\partial^2 R}{\partial p^2} = \frac{-2\beta\theta}{1 + \beta\lambda} \tag{10}$$

obtained by $\frac{\partial R}{\partial p} = 0$

So, the optimised toll is:

$$p = \frac{1}{2\beta\theta}(\alpha - \beta\mu) \tag{11}$$

substitute p into $R = q^* \times p$

and the maximum revenue is:

$$R = \frac{1}{4\beta\theta}\left[\frac{(\alpha - \beta\mu)^2}{1 + \beta\lambda}\right] \tag{12}$$

c8. Travel Demand Models

The 4-stage model can then be expressed as follows:

$$T_{ijmr} = T_i \frac{T_{ij}}{T_i} \cdot \frac{T_{ijm}}{T_{ij}} \cdot \frac{T_{ijmr}}{T_{ijm}}$$

or

$$T_{ijmr} = f_g(X_i, c_i) . f_d(T_i, c_{i\bullet}) . f_m(T_{ij}, c_{ij\bullet}) . f_r(T_{ijm}, c_{ijm\bullet})$$

where

X_i	=	vector of explanatory variables associated with zone *i*
f_g	=	trip frequency choice model **(Trip generation model)**
f_d	=	trip distribution choice model **(Trip distribution model)**
f_m	=	mode choice model **(Model split model)**
f_r	=	route choice model **(Traffic assignment model)**

One way to think about the 4-stage model is in terms of what is called a **general share** formulation. Let:

T_i	=	the number of trips produced from zone *i*
T_{ij}	=	the number of trips from zone *i* to zone *j*
T_{ijm}	=	the number of trips from zone *i* to zone *j*, using mode *m*
T_{ijmr}	=	the number of trips from zone *i* to zone *j*, using mode *m* and route *r* (which may depend on *i*, *j* and *m*)
c_i	=	denotes the cost of travel from zone *i* taking account of all available destinations, modes and routes
$c_{i\bullet}$	=	denotes the cost of travel from zone *i* to each other zone, taking account of available modes and routes
$c_{ij\bullet}$	=	denotes the cost of travel between zone *i* and zone *j* by each mode, taking account of available routes
$c_{ijm\bullet}$	=	denotes the cost of travel between zone *i* and zone *j* by mode *m* along each available route by *m* from *i* to *j*

Travel Demand Models *

The 4-Stage Model (Aggregate Approach)

Stage 1: Trip Generation (Frequency Choice Model)

Objective: Estimate number of trips produced (P_i) and attracted (A_j) by each zone

Method: Regression Models. (Establish a statistical eqn between trip ends (P_i and A_j) and zonal attributes (eg. income, car ownership)

Formula: $P_i = f(X_i, c_i, \theta) + \varepsilon$

	'To' Zone			
'From' Zone	1	2	3	
1				P_1
2				P_2
3				P_3
	A_1	A_2	A_3	

we finds all these $A_1, A_2, A_3, P_1, P_2, P_3$ in this stage.

(A_1 means total trip attracted to zone 1)
(P_1 means total trip produced from zone 1)

Stage 2: Trip Distribution (Destination Choice Model)

Objective: Determining the trip interchanges between zones, completing the Trip Matrix, T_{ij}.

Methods: Gravity Models. Interaction is proportional to activity (P_i, A_j) and inversely proportional to separation/cost (i.e. deterrence function)

doubly constrained model

Formula: $T_{ij} = P_i r_i A_j s_j f(c_{ij})$

(r_i and s_j are balancing factors solved iteratively to enforce row and column constraints. $f(c_{ij})$ is deterrence function, and typically is negative exponential: $f(c_{ij}) = e^{-\beta c_{ij}}$)

Stage 3: Mode Split (Mode Choice Model)

Objective: Split the Trip Matrix by mode (eg. car vs. bus). Find $T_{ijm} = T_{ij} \cdot Q_{ijm}$

Method: Logit Model. Find the proportion or probability of travelers using mode m, depends on the generalised cost of that mode relative to others.

Formula: $Q_{ijm} = \dfrac{T_{ijm}}{\sum_k T_{ijk}}$

$= \dfrac{f_m(c_{ijm})}{\sum_k f_k(c_{ijk})} = \dfrac{e^{-\beta c_{ijm}}}{\sum_k e^{-\beta c_{ijk}}}$

(β controls how sensitive mode choice is to cost)

Note that "$-\beta c_{ijm}$" in this case is our utility function. If given V = ... as our utility function:

$Q_{ijm} = \dfrac{e^{V_m}}{\sum_k e^{V_k}}$ (v doesn't have to be a negative function. In this case, cost c_{ijm} is a deterrence function)

Stage 4: Traffic Assignment (Route Choice Model)

Objective: Allocate Origin-Destination (OD) demand to specific routes/links on the network. (unlike stage 1-3, this stage must account for congestion, where link costs rise as flow increases)

Methods:

1. All-or-Nothing (AON) assign all traffic to the minimum cost route. (Ignore congestion)

2. Wardrop's First Principle (User Equilibrium) Traffic arranges itself so that all used routes have equal and minimum generalised (private) cost. (no individual driver can reduce the cost by switching lane)
$c(x) = c(y)$

3. Wardrop's Second Principle (System Optimal) Traffic arranges itself to minimise the Total System Cost. (Require pricing / intervention to achieve: check c8: congestion and cost part)
$T(x,y) = x c(x) + y c(y)$
$\dfrac{dT(x)}{dx} = 0$ or $\dfrac{dT(y)}{dy} = 0$

Activity-based Models (ABM) (Disaggregate Models)

Introduction

Core concept: Travel is derived from the need to participate in activities. The model schedules activities for individual rather than just counting aggregate trips.

Tours vs Trips: ABMs model Tours (chain of trips starting and ending at home) → maintaining consistency in time, space and mode.

Details: Uses high spatial (x-y coordinates) and temporal (1-minute resolution) disaggregation.

Why use ABM? *

1. Policy sensitivity: Better for analysing "soft" measures (eg. telecommuting, flex-time) and complex trade-offs.

2. Environmental precision: Can model "Cold start" vs. "Warm start" emissions based on how long a car has been parked between activities.

3. Interactions: Accounts for intra-household interactions (eg. who drops the kids at school?)

Comparison

Feature	4-Stage Model	Activity-Based Model (ABM)
Unit of Analysis	Aggregate Zones & Trips	Individual Agents & Activities/Tours
Consistency	Trips are independent (return trip not linked to outbound)	Tours ensure consistency (mode, time, location)
Data Needs	Zonal aggregate data	Disaggregate individual data
Policy Use	Infrastructure & Pricing	Soft measures, complex behaviour, detailed emissions
Weakness	Behaviourally simplistic; static	Complex to build; data intensive

Stage 1: Trip Generation Examples

Q1:

1. The following table presents data collected in the last household O-D survey (made ten years ago) for three particular zones.

Zone	Residents/HH	Workers/HH	Mean income	Population
1	3	0.9	35000	30000
2	2.5	2.2	85000	50000
3	1.6	1.2	100000	10000

Ten years ago two household-based trip generation models were estimated using these data. The first is a linear regression model given by:

$$\mathbf{y} = 0.1 + 0.6\mathbf{x_1} + 1.2\mathbf{Z_1}$$

where **y** is the mean number of trips per household per day, $\mathbf{x_1}$ is the number of workers in the household and $\mathbf{Z_1}$ is a dummy variable which takes the value of 1 for high income (>90 000) households and 0 in other cases.

The linear regression model of trip generation estimated using household (hh) O-D survey data is as follows:

$$\mathbf{y} = 0.1 + 0.6\mathbf{x_1} + 1.2\mathbf{Z_1}$$

where **y** is the mean number of trips per household per day, $\mathbf{x_1}$ is the number of workers in the household and $\mathbf{Z_1}$ is a dummy variable which takes the value of 1 for high income (>90 000) households and 0 in other cases.

Zone 1 (average workers/hh=0.9, mean hh income=35000)
y = 0.1 + 0.6 * 0.9+ 1.2 * 0 = **0.64 trips per household per day**

Zone 2 (average workers/hh =2.2, mean hh income =85000)
Y = 0.1 + 0.6 * 2.2 + 1.2 * 0 = **1.42 trips per household per day**

Zone 3 (number of workers=1.3, income=110000)
Y = 0.1 + 0.6 * 1.2 + 1.2 * 1 = **2.02 trips per household per day**

Therefore the number of trips generated in each zone can be calculated as
Number of trips per day = Trip rate **y** * Number of households
Number of trips per annum = Number of trips per day * 365

The results of these computations are presented in Table 3.

Table 3: Number of trips generated in each zone (calculated)

Zone	Number of Households	Trips per hh per day (Y)	Trips per day	Trips per annum
1	10000	0.64	6400	2336000
2	20000	1.42	28400	10366000
3	6250	2.02	12625	4608125

these are P_1, P_2 and P_3 respectively.

(should check 2025q3e)

Stage 2: Trip Distribution Examples

Q1:

Suppose we have a 3-zone system, in which the generalised cost of travel c_{ij} between zone i and zone j is given by the following matrix (in some suitable units, say pounds sterling):

	'To' Zone		
'From' Zone	1	2	3
1	0.2	1.2	2.4
2	3.8	0.2	5.7
3	0.5	3.2	0.2

this is for $f(c_{ij})$

For some future time period, a trip generation model predicts that P_i trips will be produced by zone i and A_j trips attracted to zone j. The values of P_i and A_j are as follows:

$P_1 = 400$ $\quad$ $A_1 = 400$
$P_2 = 500$ $\quad$ $A_2 = 400$
$P_3 = 300$ $\quad$ $A_3 = 400$

what we know for now is only the sum of row and column:

			400
			500
			300
400	400	400	

we want to fill this matrix!

The deterrence function is of the form $f(c_{ij}) = e^{-\beta c_{ij}}$, where $\beta = 0.02$.

(a)

1 Compute the trip matrix generated by the 'Unconstrained' gravity model:

$$T_{ij} = KP_iA_jf(c_{ij})$$

where K is a global scaling constant such that $\sum_{ij} T_{ij} = 1200$

Task 1 Unconstrained gravity model

The idea of the gravity model is that the propensity for interaction between two zones is proportional to the product of a measure of the activity in each zone divided by a measure of the separation (or travel cost) between the zones. That is:

$T_{ij} \propto P_iA_jf(c_{ij})$
and therefore
$T_{ij} = KP_iA_jf(c_{ij})$

where K is a global scaling constant such that: $\Sigma_{ij}\, T_{ij} = 1200$

first step is always to fill the trip matrix based on A_j, P_i and $f(c_{ij})$, doesn't matter what constrain model we use. (i.e. ignore s_j, r_i for now)

We first compute $T'_{ij} = P_iA_jf(c_{ij})$ as follows:
$T_{11}' = P_1 A_1 f(c_{11}) = 400 * 400 * \exp(-0.02*0.2) = 159361.278$
and so on.

$T'_{ij'} = P_iA_jf(c_{ij})$

	1	2	3	
1	159361.278	156205.714	152501.406	**468068.398**
2	185363.241	199201.598	178451.591	**563016.430**
3	118805.980	112560.600	119520.959	**350887.539**
	463530.500	**467967.911**	**450473.956**	**1381972.367**

what we do in unconstrained gravity model is make this total sum $\sum T_{ij} = A_1 + A_2 + A_3 = P_1 + P_2 + P_3 = 1200$

K is therefore calculated as $\Sigma_{i,j}\, T_{ij} / \Sigma_{i,j}\, T_{ij}' = 1200/1381972.367 = 0.003222$

we just scale the sum sum!

And the forecast trip matrix, $T_{ij} = K\, T_{ij}'$, which is calculated as follows (rounded to 0 decimal places):

T_{ij}

	1	2	3	
1	138	136	132	**406**
2	161	173	155	**489**
3	103	98	104	**305**
	402	**406**	**391**	**1200**

Note that using this method the row and column sums do not satisfy the production and attraction targets (row- and column-sums), but the total number of trips generated and attracted, i.e. $\Sigma_{ij}\, T_{ij}$, meets the target of 1200. This is the property of the unconstrained gravity model, as it is constructed to only meet the total trips constraint.

(b)

2 Compute the trip matrix generated by the 'Singly constrained' gravity model:

$$T_{ij} = P_i r_i A_j f(c_{ij})$$

where the r_i are constants such that $\sum_j T_{ij} = P_i \quad \forall i$

Task 2 Singly constrained gravity model

We are asked to apply the singly constrained gravity model with the constraint on the number of trip destinations (or attractions). That is:
$\Sigma_j T_{ij} = Pi$

This constraint leads to the following model:
$T_{ij} = P_i r_i A_j f(c_{ij})$

where r_i are constants that satisfy the constraint $\Sigma_j T_{ij} = P_i$ for all j and s_j is given by:
$r_i = [\Sigma_j A_j f(c_{ij})]^{-1}$

first step again is fill the trip matrix.

So we first calculate $T_{ij}' = P_i A_j f(c_{ij})$ as given below, i.e.
$T_{11}' = P_1 A_1 f(c_{11}) = 400 * 400 * \exp(-0.02*0.2) = 159361.278$
and so on.

$T'_{ij'} = P_i A_j f(c_{ij})$

scale row by: $\frac{400}{468068.398}$

	1	2	3	
1	159361.278	156205.714	152501.406	**468068.398**
2	185363.241	199201.598	178451.591	**563016.430**
3	118805.980	112560.600	119520.959	**350887.539**
	463530.500	**467967.911**	**450473.956**	**1381972.367**

in singly constrained model, instead scaling the TOTAL SUM, we scale the sum of row OR sum of column, based on the question.

this question wants:

$T_{ij} = P_i \underline{r_i} A_j f(c_{ij})$

we scale P_i (sum of row)

Note that this is the same as the first step in the unconstrained gravity model calculation.

The next step is to scale the elements in the T_{ij}' matrix by r_i, where
$r_1 = 1 / [A_1 * f(c_{11}) + A_2 * f(c_{12}) + A_3 * f(C_{13})] = 0.000855$
$r_2 = 1 / [A_1 * f(c_{21}) + A_2 * f(c_{22}) + A_3 * f(C_{23})] = 0.000888$
$r_3 = 1 / [A_1 * f(c_{31}) + A_2 * f(c_{32}) + A_3 * f(C_{33})] = 0.000855$

and $T_{11} = T_{11}' * r_1$ and so on.

The forecast trip matrix is therefore as follows

T_{ij}

	1	2	3	
1	136.186	133.490	130.324	**400.000**
2	164.616	176.906	158.478	**500.000**
3	101.576	96.236	102.187	**300.000**
	402.379	**406.632**	**390.990**	**1200.000**

The final forecast trip matrix rounded up to 0 decimal places is as follows.

T_{ij}

	1	2	3	
1	136	133	130	**400**
2	165	177	158	**500**
3	102	96	102	**300**
	402	**407**	**391**	**1200**

Note that using this method we satisfy the row constraints i.e. the number of trips generated from a zone, by design. But we do not meet the column targets of trip attractions.

(c)

3 Compute the trip matrix generated by the 'Doubly constrained' gravity model:

$$T_{ij} = P_i r_i A_j s_j f(c_{ij})$$

where the r_i are constants such that $\sum_j T_{ij} = P_i \;\; \forall i$ and s_j are constants such that $\sum_i T_{ij} = A_j \;\; \forall j$

Task 3 Doubly constrained gravity model

It's just iterating singly constrain model again and again, from scaling row first, then scaling column, then back to scaling row and so on, until it converges.

The doubly constrained gravity model is designed to meet both the production and attraction constraints i.e. $\Sigma_i T_{ij} = A_j$ and $\Sigma_j T_{ij} = P_i$

This leads to the model:
$T_{ij} = P_i r_i A_j s_j f(c_{ij})$

where
r_i constants that satisfy the constraint $\Sigma_j T_{ij} = P_i$ for all i
s_j constants that satisfy the constraint $\Sigma_i T_{ij} = A_j$ for all j

These are given by:
$r_i = [\Sigma_j A_j s_j f(c_{ij})]^{-1}$
$s_j = [\Sigma_i P_i r_i f(c_{ij})]^{-1}$

Since r_i is a function of s_j and vice versa, this does not have a direct solution and can be solved only through an iterative method.

ITERATION 1

In iteration 1 we are required to meet the production constraints i.e. $\Sigma_j T_{ij} = P_i$
So we start with the convenient assumption that s_j=1 for all j, which yields the following model
$T_{ij} = P_i r_i A_j f(c_{ij})$
where
$r_i = [\Sigma_i A_j f(c_{ij})]^{-1}$

This is the same as the singly unconstrained model that meets attraction constrains. Therefore at the end of iteration 1 we have the trip matrix from part (b)

this is after scaling row.

$T_{ij}(1)$		1	2	3	
	1	136.186	133.490	130.324	**400.000**
iteration 2:	2	164.616	176.906	158.478	**500.000**
	3	101.576	96.236	102.187	**300.000**
		402.379	**406.632**	**390.990**	**1200.000**
		scale column by $\frac{400}{402.379}$	scale column by $\frac{400}{406.632}$	scale column by $\frac{400}{390.990}$	

ITERATION 2

In iteration 2, we use the output of iteration 1, i.e. $T_{ij}(1)$, and calculate $T_{ij}(2)$ that meets the attraction constraints i.e. $\Sigma_i T_{ij} = A_j$. That is:
$T_{ij}(2) = A_j * [T_{ij}(1)/ \Sigma_k T_{kj}(1)]$, which computes as

$T_{11}(2) = A_1 * \{T_{11}(1)/[T_{11}(1)+T_{21}(1)+T_{31}(1)]\}$
$= 400 * [136.186/ (136.186+164.616+101.576)] = 135.381$
and so on.

The result of iteration 2 is as follows

$T_{ij}(2)$		1	2	3	
	1	135.381	131.313	133.327	**400.021**
	2	163.643	174.021	162.130	**499.794**
	3	100.976	94.667	104.542	**300.185**
		400.000	**400.000**	**400.000**	**1200.000**

ITERATION 3

In iteration 3, we use the output of iteration 2, i.e. $T_{ij}(2)$, and calculate $T_{ij}(3)$ that again meets the production constraints i.e. $\Sigma_j T_{ij} = P_i$. That is:
$T_{ij}(3) = P_i * [T_{ij}(2)/ \Sigma_k T_{ik}(2)]$, which computes as

$T_{11}(3) = P_1 * \{T_{11}(2)/[T_{11}(2)+T_{12}(2)+T_{13}(2)]\}$
$= 400* [135.381/(135.381+131.313+133.327)] = 135.374$
and so on.

The result of iteration 3 is as follows:

		1	**2**	**3**	
	1	135.374	131.306	133.320	**400.000**
$T_{ij}(3)$	**2**	163.711	174.092	162.197	**500.000**
	3	100.913	94.609	104.478	**300.000**
		399.998	**400.006**	**399.995**	**1200.000**

ITERATION 4

Finally, in iteration 4, we use the output of iteration 3, i.e. $T_{ij}(3)$, and calculate $T_{ij}(4)$ that meets the attraction constraints i.e. $\Sigma_i T_{ij} = A_j$. That is:
$T_{ij}(4) = A_j * [T_{ij}(3)/ \Sigma_k T_{kj}(3)]$ which computes as:

$T_{11}(4) = A_1 * \{T_{11}(1)/[T_{11}(1)+T_{21}(1)+T_{31}(1)]\}$
$= 400 * [135.374/ (135.374+163.711+100.913)] = 135.375$
and so on.

The result of iteration 4 is as follows

		1	**2**	**3**	
	1	135.375	131.303	133.322	**400.000**
$T_{ij}(4)$	**2**	163.711	174.089	162.199	**500.000**
	3	100.914	94.607	104.479	**300.000**
		400.000	**400.000**	**400.000**	**1200.000**

The final trip matrix using the doubly constrained model, and rounded up to 0 decimal places is

		1	**2**	**3**	
	1	135	131	133	**400**
T_{ij}-final	**2**	164	174	162	**500**
	3	101	95	104	**300**
		400	**400**	**400**	**1200**

During each of the iterations the trip matrix satisfies one of either the row or column constraints exactly. And as seen from the outputs above, with each iteration the results come closer to convergence i.e. to satisfying both the row and column constraints simultaneously. At the end of iteration 4, we almost satisfy both constrain simultaneously while at the same time taking into consideration the cost of travel between each zone.

Stage 3 : Mode Split Example

Q3.

A binomial logit model has been calibrated and the following utility functions obtained:

$V_c = 0.4 - 0.40t_i - 0.90t_o - 0.032c_c - 0.08c_p + 1.2M + 0.05I$

$V_r = -0.35t_i - 0.95t_o - 0.08c_r$

where V_c is the utility function for the car mode
V_r is the utility function for the rail mode
t_i is the in-vehicle time in minutes
t_o is the out-of-vehicle time in minutes
c_c is the cost of travel by car in pence
c_p is the cost of car parking in pence
c_r is the cost of travel by rail in pence
M is a dummy mode that takes a value of 0 for males and 1 for females
I is the income in £1000s per year

1. Interpret the value of the parameters in the utility functions above.

2. For a particular man with an income of £50,000 per year, the variables have the following values:

	Car	Rail
In-vehicle time in minutes	15	12
Out-of-vehicle time in minutes	5	10
Cost of travel in pence	80	120
Cost of parking in pence	150	-

(a) What would be the probability of him travelling by car?

(b) What would be the probability of a woman with the same salary travelling by car?

(c) What would the probabilities of travelling by car for each of them change to if the cost of parking was increased to 180 pence?

2.
(a) For a particular man (i.e. $M = 1$) with

Annual income = I = £50,000 = 50 (in £1000s)
In-vehicle travel time; t_i (car) = 15 min, t_i (rail) = 12 min
Out-of-vehicle travel time; t_o (car) = 5 min, t_o (rail) = 10 min
Cost of travel by car = c_c = 80 pence
Cost of travel by rail = c_r = 120 pence
Cost of parking = c_p = 150 pence

Therefore:
$V_{car} = 0.4 - 0.40t_i - 0.90t_o - 0.032c_c - 0.08c_p + 1.2M + 0.05I$
$= 0.4 - 0.40*15 - 0.90*5 - 0.032*80 - 0.08*150 + 1.2*1 + 0.05*50$
$= -20.960$
$V_{rail} = -0.35t_i - 0.95t_o - 0.08c_r$
$= -0.35*12 - 0.95*10 - 0.08*120$
$= -23.300$
Probability of the man travelling by car = $\exp(V_{car})/[\exp(V_{car}) + \exp(V_{rail})] = 0.912$

(b) The probability of a woman in the same circumstances travelling by car is calculated as follows:

All the variables are the same except M=0
$V_{car} = -22.160$
$V_{rail} = -23.300$

Probability of the woman travelling by car = $\exp(V_{car})/[\exp(V_{car}) + \exp(V_{rail})] = 0.758$

In other words, all else being equal, this model suggests that a woman is less likely to travel by car than a man. This is due to the negative gender (male) variable in the disutility for rail, suggesting that men inherently assign a higher disutility to travelling by rail than women.

(c) Now, if the cost of parking was increased to 180 pence i.e. c_p = 180 pence

For the man (i.e. $M = 1$)
$V_{car} = 0.4 - 0.40t_i - 0.90t_o - 0.032c_c - 0.08c_p + 1.2M + 0.05I$
$= 0.4 - 0.4*15 - 0.90*5 - 0.032*80 - 0.08*180 + 1.2*1 + 0.05*50$
$= -23.360$
$V_{rail} = -0.35t_i - 0.95t_o - 0.08c_r$
$= -0.35*12 - 0.95*10 - 0.08*120$
$= -23.300$

Probability of the man travelling by car = $\exp(V_{car})/[\exp(V_{car}) + \exp(V_{rail})] = 0.485$

For the woman (i.e. $M = 0$)

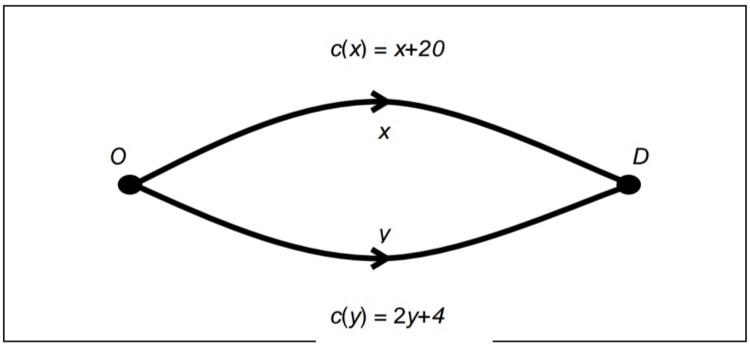
$V_{car} = 0.4 - 0.40t_i - 0.90t_o - 0.032c_c - 0.08c_p + 1.2M + 0.05I$
$= 0.4 - 0.4*15 - 0.90*5 - 0.032*80 - 0.08*180 + 1.2*0 + 0.05*50$
$= -24.560$

$V_{rail} = -0.35t_i - 0.95t_o - 0.08c_r$
$= -0.35*12 - 0.95*10 - 0.08*120$
$= -23.300$

Probability of the man travelling by car = $\exp(V_{car})/[\exp(V_{car}) + \exp(V_{rail})] = 0.221$

Stage 4 : Traffic Assignment Model

Q1. Suppose we have a system comprising a single origin O connected to a single destination D via a congestable network consisting of two disjoint routes, each represented by a single link, as shown below.



The demand for travel between O and D is q and this demand induces flows of x on the upper route and y on the lower route.

The generalised (private) cost of using the upper route is $c(x) = x+20$ and generalised (private) cost of using the lower route is $c(y) = 2y+4$.

The demand q between O and D is fixed at 80.

TASKS

1 Compute the traffic assignment corresponding to a Wardropian (WP1) equilibrium in this network and hence find the equilibrium travel time between O and D and the total system cost.

2 In what ways would you expect a system optimising (WP2) flow pattern in this network to differ from the WP equilibrium?

SOLUTION

Task 1 (WP1's example) — $c(x) = c(y)$

The solution to the Wardropian equilibrium problem (WP1) is that assignment which results in equal journey time along each route.

Therefore at a WP1 equilibrium

$$x+20 = 2y+4 \qquad (1)$$

and by conservation of flow we also have

$$x+y=80 \qquad (2)$$

Equations (1) and (2) are a pair of linear simultaneous equations in two variables, and are easily solved by direct substitution (e.g. for y in (1) from (2)). This gives,

$$x+20 = 2(80-x)+4$$
$$\Rightarrow x = 48, \text{ and } y = 32.$$

So in the WP1 solution, the cost along each routue is 68 and the total network cost TC is $48*68 + 32*68 = 80*68$. Hence $TCWP1 = 5440$.

Task 2 (WP2's example) — $T = xc(x) + yc(y)$ $\frac{\partial T}{\partial x} = 0$ or $\frac{\partial T}{\partial y} = 0$

An assignment according to Wardrop's Second Principle is one in which the total network cost is minimised.
For a given assignment, the total network cost *TC* is

$$TC = x(x+20) + y(2y+4) \tag{3}$$

Using the conservation equation (2), we can express *TC* in terms of either *x* or *y* alone.
Expressed in terms of *x*,

$$TC(x) = x(x+20) + (80-x)[2(80-x)+4]$$

Hence

$$TC(x) = 3x^2 - 304x + 13120 \tag{4}$$

TC is minimised when $\frac{\partial TC(x)}{\partial x} = 0$ and $\frac{\partial^2 TC(x)}{\partial x^2} > 0$

$\frac{\partial TC(x)}{\partial x} = 6x - 304$ and $\frac{\partial^2 TC(x)}{\partial x^2} = 6\ (>0)$

Hence $x = 50.7$ and $y = 29.3$ is a minimum.

Notice that in the WP2 solution the costs along each route are not equal. Along the upper route the cost is 70.7 and along the lower route it is 62.6. Hence, relative to the WP1 equilibrium, travellers on the upper route are worse off but travellers along the lower route are better off.

Substituting into (4) gives *TCWP2* = 5418.7.

As we would expect, *TCWP1* > *TCWP2*.
Hence, despite the fact that in a WP1 solution each individual seeks to achieve the best possible outcome for themselves, the aggregate outcome for everybody is not optimal. The reason for this difference is the divergence between individual (private) costs and marginal (social) costs. Intuitively, what happens is that in the attempt to achieve the best possible outcome for themselves each driver interferes with the passage of other drivers, and does so more than is absolutely necessary. When all drivers behave in this way the system is in aggregate worse off.

* the differences in x and y for WP1 and WP2 are the "blind user" who should choose WP2 over WP1 but didn't because of selfishness

In this case in the WP1 solution, too many drivers (=50.7-48=2.7) choose to use the lower route and although these particular individuals are better off using this route than the upper route, they cause extra congestion for themselves and all the other users on the lower route.

c9. Transport Environment

$$\text{Impact} = \text{Activity} \times \frac{\text{Energy}}{\text{Activity}} \times \frac{\text{Emissions}}{\text{Energy}} \times \frac{\text{Damage}}{\text{Emissions}}$$

Avoid / Reduce
- Teleconferencing instead of flying

Shift / Improve (Reduce F)
- shift from car to rail
- improve car's aerodynamic

Improve
- Shift from diesel to electric

Better planning
- Keep high polluting vehicle from population center
- Taller chimney stacks for dispersion

Impact: e.g. Climate or air quality

Activity: e.g. km, pkm, vkm, tonne-km

Energy intensity:
- e.g. kWh/km, kWh/pkm, kWh/vkm or kWh/tonne-km
- Depends on:
  - Mode
  - Powertrain
  - Fuel
  - Duty cycle
  - Passenger occupancy

full bus' I < empty bus' I

Emissions factor (energy):
- e.g. kg CO_2/kWh
- Depends on:
  - Mode
  - Powertrain
  - Fuel
  - Duty cycle
  - Emissions control

Damage factor:
- Depends on:
  - Species
  - Location
  - Source
  - Population
- Can be a monetised damage cost
  - e.g. £/kg

Transport Environment

Transport Energy

1. Key Physics and Equations

$F_{drag} = \frac{1}{2}\rho A_f C_D V^2$ (Aerodynamic drag)

air density, frontal area, drag coeff., speed relative to air

$F_{roll} = C_R m g$ (Rolling resistance)

coeff of rolling resistance (depends on tyre/road only!), mass of vehicle, gravity

$F_{grad} = mg\sin\theta$ (Gradient Force)

Weight of the vehicle in the slope direction.

$(\underbrace{F_{propulsion}}_{\text{can be -ive (eg. going downhill, } a=0)} - F_{drag} - F_{roll} - F_{grad} = ma)$

2. Work and Power

Work (Energy) = Force × Distance (propulsion force.)

(unit: J or kWh) (if force varies over x, $F(x)$: $W = \int_0^x F(x)\,dx$)

Power (Rate of Energy Consumption)

= $F_{total} \times V$ (velocity)

(note: $F_{drag} \propto V^2$, hence the power required to overcome drag is proportional to V^3, hence speed limit saves fuel)

3. Energy Intensity (I=AIED)

- used to compare energy efficiency across different mode (eg. Rail vs Car)

- Energy Intensity (unit: kWh/km, kWh/pkm, kWh/vkm) $= \dfrac{\text{Energy consumed (unit kWh)}}{\text{Activity (unit: km, pkm, vkm)}}$

$= \dfrac{\text{Force} \times \text{distance}}{\text{distance}}$

$=$ Force ✱ (if pkm ÷ by p, if vkm ÷ by v)

Impact Pathway

- Transport activity
- Energy consumption
- Emissions (e.g. kg of GHG or other pollutants)
- Atmospheric dispersion and processing
- Impact and response (e.g. climate, human health)
- Monetary evaluation of costs (£££)

(Energy consumption, Emissions: combustions)

Combustion and Emissions

1. Combustion

Ideal: Fuel (C_xH_y) + Air → CO_2 + H_2O + Heat

$$C_xH_y + \left(x+\frac{y}{4}\right)O_2 \rightarrow xCO_2 + \frac{y}{2}H_2O$$

Real: Usually incomplete combustion, which leads to by product such as:

- CO (carbon monoxide): incomplete oxidation
- UHC (unburned hydrocarbon)
- PM (particulate matter): soot/black carbon (incomplete oxidation) + non-exhaust sources (brake/tyre wear)
- NO_x (nitrogen oxide): Formed at high temp. when N_2 in air oxidises (NO nitric oxide, NO_2 nitrogen dioxide)

2. Greenhouse Gases (GHGs) vs. Air Pollutants

GHGs (Global Impact): Causes climate change

- primary CO_2, CH_4, N_2O
- Global Warming Potential (GWP) is a metric to compare GHGs to CO_2

$CO_2e = \sum(\text{Mass}_i \times GWP_i)$

eg. CO_2e of CO_2 = 1, CO_2e of CH_4 ≈ 28 (over 100yrs)

GWP depends on timeframe!

3. Emission Factors (I=AIED)

- Can be energy based form or distance based form

$$\text{Impact} = \text{Activity} \times \frac{\text{Energy}}{\text{Activity}} \times \underbrace{\frac{\text{Emissions}}{\text{Energy}}} \times \frac{\text{Damage}}{\text{Emissions}}$$

energy based form:

emission factor $= \dfrac{\text{emission}}{\text{energy}}$ (unit: kgCO_2/kWh)

(I=AIED)

$$\text{Impact} = \text{Activity} \times \underbrace{\frac{\text{Energy}}{\text{Activity}} \times \frac{\text{Emissions}}{\text{Energy}}} \times \frac{\text{Damage}}{\text{Emissions}}$$

distance based form:

emission factor $= \dfrac{\text{emission}}{\text{activity}}$ (unit: kgCO_2/km)

Pollution Dispersion

1. Simple Box Model

We assume that the street is a "box". We want to solve for steady state concentration (χ)

Formula: $\chi = \chi_B + \dfrac{L \times q}{U \times H}$

- χ: conc. in the street (μg/m³ or kg/m³)
- χ_B: background conc.
- L: length of the street.
- H: mixing height (often taken as height of building)
- U: wind speed parallel to the st.
- q: emission rate per unit area (kg/m²s)

2. Calculating 'q'

How to find 'q' emission rate per area? (from traffic flow)

$$q = \frac{Q \times E}{W}$$

- Q: traffic flow (veh/s)
- E: emission factor per veh per m (g/m or kg/m)
  - → note: if E is given in g/km, convert to g/m by dividing by 1000.
- W: width of the street.

How to Remember Implicit Form Mathematical Model? "SPVOC"

1. Sets ("Who" and "Where")

Lists the things involved. Look for nouns.

eg. Are there sources and destination? let S be sets of node of supply, D be sets of node of demand. (in some cases we treat them as subsets of N.)

Are there nodes and link? let N be set of nodes, E be set of links.

Are there different commodities? let P be set of commodities.

2. Parameters (the "Known")

Data given in the table (costs, capacities, supplies, demands)

eg. cost to move from i to j → c_{ij}

distance from i to j → d_{ij}

capacity of link (i,j) → L_{ij}

demand or supply at node i → A_i or B_i (or if generalised like $S, D \in N$: we call it just D_i)

3. Variables (the "Decision")

What we are after for. Ask "how much" or "Yes/No" questions.

eg. continuous (x_{ij}) → How much flow of goods from i to j?

binary (y_j) → Is facility j opened?

4. Objective (the "Goal")

It is usually minimise cost, maximise profit or minimise distance travelled.

eg. $\min \sum_{(i,j)\in E} c_{ij}x_{ij}$ or $\min \sum_{j\in D}\sum_{i\in S} c_{ij}x_{ij}$ (add $\sum_{i\in S} C_i y_i$ for fixed opening cost in C-FLP)

5. Constraints (HARDEST)

supply, demand, capacity, nature. } for FLP is a bit special

Family 1: Supply / Demand

eg. $\underbrace{\sum_{j\in D} x_{ij} = A_i, \forall i\in S}_{\text{supply constraints}}$, $\underbrace{\sum_{i\in S} x_{ij} = B_i, \forall j\in D}_{\text{demand constraints}}$ (FLP (single or multi): $\sum_{i\in S} x_{ij} = 1, \forall j\in D$)

* no change... $= B_i$ if capacitated.

if combined:

$$\underbrace{\sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = D_i}_{\text{flow conservation constraints.}}, \forall i\in N \quad \left(\text{shortest path: } D_i = \begin{cases} 1, & i=s \\ -1, & i=t \\ 0, & \text{otherwise} \end{cases}\right)$$

Family 2: Capacity Constraints.

eg. $x_{ij} \le L_{ij}, \forall (i,j)\in E$

($\sum_{c\in P} x_{ij}^c \le L_{ij}$ if multi-commodity)

Family 3: The FLP Supply Constraints.

(demand already shown in family 1)

1. no. of facilities (supply) opened.

$\sum_{i\in S} y_i = 1$ (single), $\sum_{i\in S} y_i = n$ (multiple)

$n_{min} < \sum_{i\in S} y_i < n_{max}$ (capacitated)

2. capacity of supply node i (if open)

$x_{ij} \le y_i, \forall i\in S, \forall j\in D$ (single FLP, multi FLP)

$\sum_{j\in D} x_{ij} \le P_i y_i, \forall i\in S$ (capacitated FLP)

Family 4: Number nature constraints.

eg. $x_{ij} \ge 0, \forall (i,j)\in E$ (flow on links are non negative)

$y_i \in [0,1], \forall i\in S$ (supply facility node open or close)