CITY UNIVERSITY OF HONG KONG

Course code & title : EE 3301 Optimization Methods for Engineering
Session : Semester A 2021/22
Time allowed : 2 hours (plus 20 minutes only for uploading)

- This paper has **<u>SEVEN</u>** pages (including this cover page).
- Online questions are not to be distributed or retained in any form after the exam.

Instructions

Please make sure you follow all instructions from the University, ARRO, and EE. Please note the following:

1. This paper consists of **<u>FIVE</u>** questions that have multiple sub-questions. ALL the questions and sub-questions are compulsory. Make sure that you attempt all of them. The total score is 100.
2. This is an **open-book exam.** Students can read the lecture notes and/or other materials available online.
3. You are responsible for receiving the questions on Canvas. Hand-write* all answers in the answer-book or on blank answer sheets, compile the answers into a single PDF file and **submit the PDF file and three Excel files to the Canvas Assignment called "Final Exam" before the exam deadline separately**, **NOT** in a Zip file. See additional instructions on Page 2.
4. **Stay in your seat after the deadline** until the examiner allows you to leave.
5. You were given a 10-digit code to be used in questions as instructed. You must use only the new code. See additional instructions on Page 2.

* You are allowed to copy and paste text or figures and to fill in/replace numbers by typing.

You must cite the source if the material is not part of the exam questions.

Answering this exam paper implies your acknowledgment of the Pledge for following the Rules on Academic Honesty:

"I pledge that the answers in this examination are my own and that I will not seek or obtain an unfair advantage in producing these answers. Specifically,

1. I will not plagiarize (copy without citation) from any source;
2. I will not communicate or attempt to communicate with any other person during the examination; neither will I give or attempt to give assistance to another student taking the examination; and
3. I will use only approved devices (e.g., calculators) and/or approved device models.
4. I understand that any act of academic dishonesty can lead to disciplinary action."

On the first page of your answer sheets, copy the following sentence and sign it: *I pledge to follow the Rules on Academic Honesty and understand that violations may lead to severe penalties.*

(Signature) ____________________ (Date) ________________________

<u>Additional Instructions</u>

You **MUST** use the 10-digit code that you received **in the Module called Final Exam.** Do **NOT** use any other sequence of numbers as your 10-digit number. Using the wrong number will incur **PENALTY** in grading.

You must submit **Excel files** as instructed. If you submit solutions using other software by other vendors, there may be **PENALTY** in grading. Note that in Group Learning, you are encouraged to use other software tools, but in the exam, you must follow the instructions.

Should you have any technical problems during the exam, contact your course leader or invigilator.

<u>Question 1 (32 marks)</u>

Answer the following True or False questions. You must provide a clear and rigorous justification for all your answers.

1.1. (3 marks) Consider the map of Konigsberg in the 18th century, as shown in Figure 1. It is possible to find a walk through the city that crosses each bridge exactly once. True/False

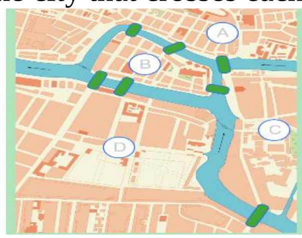


Figure 1

False.

Consider each location among (A, B, C, D) as a node and each bridge as an edge. Therefore, from the map, the degrees of nodes are as follows,

- Location A has degree = 3
- Location B has degree = 5
- Location C has degree = 3
- Location D has degree = 3

There are more than two nodes/locations with an odd degree, so it is impossible to find a walk through the city that crosses each bridge exactly once.

1.2. (3 marks) Consider today's map of Konigsberg shown in Figure 2. It is possible to find a walk through the city that crosses each bridge exactly once. True/False

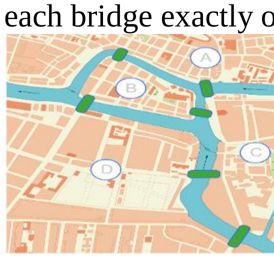


Figure 2

True.

Consider each location among (A, B, C, D) as a node and each bridge is an edge. Therefore, from the map, the degrees of nodes are as follows,

- Location A has degree = 2
- Location B has degree = 3
- Location C has degree = 4
- Location D has degree = 3

There are only two nodes/locations with an odd degree (B, D) so, it is possible to find a walk through the city that crosses each bridge exactly once. One of the two locations (B, D) must be the start point and the other must be the endpoint of the walk.

An example for such a walk is (B-A-C-B-D-C-D).

1.3. (3 marks) The sum of all the entries of an incidence matrix of an unweighted directed graph is zero. True/False

True.
In the case of an unweighted directed graph, each link between two nodes is represented by a column in the incidence matrix; each column contains two values (-1,+1) where -1 means the edge leaves the node and +1 means the edge enters the node. Therefore, the summation of each column entry is equal to zero. So the sum of all the entries of an incidence matrix of an unweighted directed graph equals zero.

1.4. (2 marks) For a simple, undirected, and unweighted graph, the diagonal entries of the adjacency matrix are always zeros. True/False

True.
The diagonal entries of the adjacency matrix represent self-loops. A simple, undirected, and unweighted graph does not have self-loops according to its definition. Therefore, the diagonal entries of its adjacency matrix are zeros.

1.5. (3 marks) The diameter of a simple, connected, undirected, and unweighted graph that has 26 nodes, can be greater than 25. True/False

False.
The diameter is the maximum legths of all shortest paths between any pairs of nodes. Therefore, among all possible connected graphs with 26 nodes, the largest possible diameter when this graph is connecting as a “chain” of 26 nodes, in this case, the largest possible diameter will be (26-1=25).

1.6. (3 marks) It is possible for a simple, connected, undirected, and unweighted graph that has 26 nodes to have a diameter of 2. True/False

True.
For a star topology, the shortest path between any two node in the periphery is two links – one to the center-node and one from the center node to the other node. And the shortest path beteen the center node and any other node is one link. Therefore the maximum shortest path is 2, so the diameter is equal to 2.

Note: For the unweighted graphs of sub-questions 1.5 and 1.6, assume that each edge has a weight of 1 for the purpose of diameter calculation.

1.7. (2 marks) For a simple graph that has an odd number of nodes, the total degree is always an even number. True/False

True.
The total degree in a simple graph must be an even number because each link contribute 2 to the total degree (1 for each end-node) so the total degree must be 2|E| which is even.

1.8. (2 marks) A simple, undirected, and unweighted graph with eight nodes cannot have more than 28 edges. True/False

True.
For a simple, undirected, and unweighted graph of eight nodes, the largest possible number of edges if it is a fully connected/complete graph. In this case, the largest possible number of edges is $\frac{n*(n-1)}{2} = \frac{8*7}{2} = 28$ edges.

1.9. (3 marks) Consider the problem

$$\text{Maximize } (x+8)^2 + 5 + e^x$$
$$\text{such that}$$
$$6 \geq x \geq 0.$$

This problem is a convex optimization problem. True/False

False.
As the objective function is convex, it must be minimized subject to convex feasible region in a convex optimization problem. Here the objective function is maximized, so the answer is False.

1.10. (3 marks) For a directed graph, the adjacency matrix must be asymmetric. True/False

False.
The adjacency matrix of a directed graph may be symmetric if each pair of connected nodes have two connections in opposite directions, such as (an edge from node A to B and an edge from node B to A).

1.11 (2 marks) Let $f_1(\mathrm{x}) = -\,\mathrm{e}^{\mathrm{x}}$, $f_2(\mathrm{x}) = -\mathrm{x}^3$, $f_3(\mathrm{x}) = f_1(\mathrm{x}) + f_2(\mathrm{x})$. The function $f_3(\mathrm{x})$ must be a concave function. True/False

False.
Because $f_1(\mathrm{x}) = -\,\mathrm{e}^{\mathrm{x}}$ is a concave function, and $f_2(\mathrm{x}) = -\mathrm{x}^3$ is not a concave function, therefore the summation of $f_1(\mathrm{x})$ *and* $f_2(\mathrm{x})$ is not a concave function.

1.12 (3 marks) Consider a fully connected network where every node has a direct link to any other node. Such a network will always maintain connectivity among the nodes for any link failure event. True/False

False
If the number of nodes is 2, there is only one link and if it fails the network is disconnected.

<u>Question 2 (20 marks)</u>
Consider a traffic monitoring system. The traffic data is shown below:

Point	1	2	3	4	5	6	7	8	9	10

Day	1	3	5	7	9	11	12	14	16	18
Number of Users	6 + *N1*	18 + *N2*	29 + *N3*	47 + *N4*	59 + *N5*	73 + *N6*	85 + *N7*	97 + *N8*	120 + *N9*	142 + *N10*

For example, if your 10-digit number is 5494433035, the table should be updated to:

Point	1	2	3	4	5	6	7	8	9	10
Day	1	3	5	7	9	11	12	14	16	18
Number of Users	6 + 5 = 11	18 + 4 = 22	29 + 9 = 38	47 + 4 = 51	59 + 4 = 63	73 + 3 = 76	85 + 3 = 88	97 + 0 = 97	120 + 3 = 123	142 + 5 = 147

Point	1	2	3	4	5	6	7	8	9	10
Day	1	3	5	7	9	11	12	14	16	18
Number of Users	11	22	38	51	63	76	88	97	123	147

2.1. (1 mark) Write down your 10-digit number.

2.2. (1 mark) Update the table using your specific *Ni* value in the number of users in the ith point.

Remark: *The value Ni is the ith digit of your 10-digit number.*
For example, if your 10-digit number is 5494433035, the table should be updated to:

Point	1	2	3	4	5	6	7	8	9	10
Day	1	3	5	7	9	11	12	14	16	18
Number of Users	6 + 5 = 11	18 + 4 = 22	29 + 9 = 38	47 + 4 = 51	59 + 4 = 63	73 + 3 = 76	85 + 3 = 88	97 + 0 = 97	120 + 3 = 123	142 + 5 = 147

Point	1	2	3	4	5	6	7	8	9	10
Day	1	3	5	7	9	11	12	14	16	18
Number of Users	11	22	38	51	63	76	88	97	123	147

2.3. (8 marks) The aim is to introduce a linear regression using a linear function in the form of $y = ax + b$ to fit these data. Provide a convex optimization formulation for this regression with an objective to find the optimal values for a and b to minimize the sum of the squares of the residuals.

Find the optimal values for a and b that minimize:

$$S = \sum_{i=1}^{10}(y_i - ax_i - b)^2, i = 1, 2, 3, \dots, 10$$

where the values of x_i and y_i are given below

Point	1	2	3	4	5	6	7	8	9	10
x_i	1	3	5	7	9	11	12	14	16	18
y_i	11	22	38	51	63	76	88	97	123	147

2.4. (10 marks) Solve the convex optimization problem formulated in sub-question 2.3 in two ways: (1) using Excel Solver and (2) using the analytical solution. After verifying consistency between (1) and (2), use the optimal values of a and b to predict the number of users on the 17th day.

Remark: *Name the Excel file* "Q2_Regression" *and upload it with your answers on Canvas.*

Start with the initial solution $\hat{a} = 0$ and $\hat{b} = 0$. Using Excel solver, we obtain the optimal values for a and b:

$$\hat{a} = 7.67017$$
$$\hat{b} = -2.03359$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Convex Optimization														
2				y-intercept	b=	-2.03359	(Intercept)	Decision							
3				slope	a=	7.67017	(slope)	variables						n=	10
4															
5															
6				x	x	1	3	5	7	9	11	12	14	16	18
7															
8				y	y	11	22	38	51	63	76	88	97	123	147

For the analytic solution, we define $\bar{x}$, the average of all x_i as:

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10}$$

and $\bar{y}$, the average of all y_i as:

$$\bar{y} = \frac{\sum_{i=1}^{10} y_i}{10}$$

Then

$$\hat{a} = \frac{\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10}(x_i - \bar{x})^2} = \frac{10\sum_{i=1}^{10} x_i y_i - \sum_{i=1}^{10} x_i \sum_{i=1}^{10} y_i}{10\sum_{i=1}^{10} {x_i}^2 - \left(\sum_{i=1}^{10} x_i\right)^2},$$

and

$$\hat{b} = \bar{y} - \hat{a}\bar{x}.$$

The result of the analytical solution is:

$$\hat{a} = 7.670183$$
$$\hat{b} = -2.03376$$

We can see that the results of these two approaches are very close to each other.

The number of users on the 17th day is approximately equal to $\hat{a} \times 17 + \hat{b}$ = 128.3593.

Question 3 (20 marks)

Consider a packet switching network whose topology is the directed graph shown in Figure 3. In the graph, each unidirectional link is associated with an integer number representing its capacity, i.e., the maximum data rate carried on this link. Node 1 and Node 4 have a server as well as an optical switch (without buffering), while each of the other nodes only has an optical switch. The server at Node 1 stores a large amount of data that must be sent urgently to the server at Node 4 It is therefore important to send the data from Node 1 to Node 4 at a maximum transmission rate. However, we must consider the capacity limitations of the links and also the flow conservation constraints.

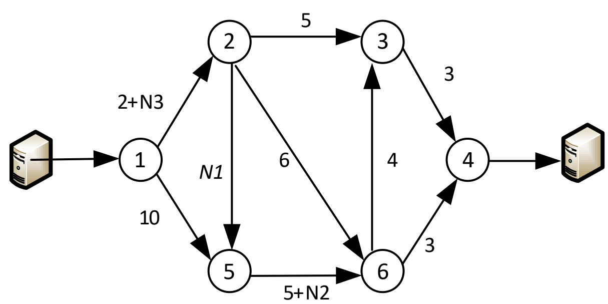


Figure 3

Remarks:

1. *The value Ni is the ith digit of your 10-digit number.*
2. *The capacities of the local links from the server of Node 1 to Node 1 and from Node 4 to the server of Node 4 are assumed to be much larger than the capacities of the links between the different nodes, so these local links can assumed to have unlimited capacities. Therefore, the aim is to maximize the flow between Node 1 and Node 4.*

3.1. (1 mark) Write down the capacities of links (1,2), (2,5), and (5,6).

(Assume *N1* = 1, *N2* = 0, *N3* = 9.)
The capacities of links (1,2), (2,5), and (5,6).
are 2 + *N3* = 11, *N1* = 1, and 5 + *N2* = 5, respectively.

3.2. (7 marks) Provide a linear programming formulation to this problem of sending the data from Node 1 to Node 4 at the maximum transmission rate. Clearly define the variables and provide an explanation of the objective function. List all the constraints and explain these constraints.

Define the decision variable as follows.

x_{ij} = The rate of data carried on link (i, j),

= *or* the flow that represents the data rate allocated by link (i, j),

= *or* the traffic on link (i, j).

Objective: Maximize the total amount of flow egressing Node 1, (or the total amount of flow ingressing Node 4), which is formulated as:

$$\max(x_{12} + x_{15}) \text{ or } \max(x_{34} + x_{64})$$

Constraints: There are mainly two types of constraints: capacity constraints and flow-conservation constraints:

(1) Capacity constraints: The carried data rate on each link cannot surpass the link capacity.

$$x_{12} \leq 2 + N3 = 11,$$
$$x_{15} \leq 10,$$
$$x_{23} \leq 5,$$
$$x_{25} \leq N1 = 1,$$
$$x_{26} \leq 6,$$
$$x_{34} \leq 3,$$
$$x_{56} \leq 5 + N2 = 5,$$
$$x_{63} \leq 4,$$
$$x_{64} \leq 3.$$

(2) Flow conservation constraints, i.e., the flow egressing an intermediate node is equal to the flow entering this node.

$$\text{Node2}: x_{12} = x_{23} + x_{25} + x_{26},$$
$$\text{Node3}: x_{23} + x_{63} = x_{34},$$
$$\text{Node5}: x_{15} + x_{25} = x_{56},$$
$$\text{Node6}: x_{26} + x_{56} = x_{63} + x_{64}.$$

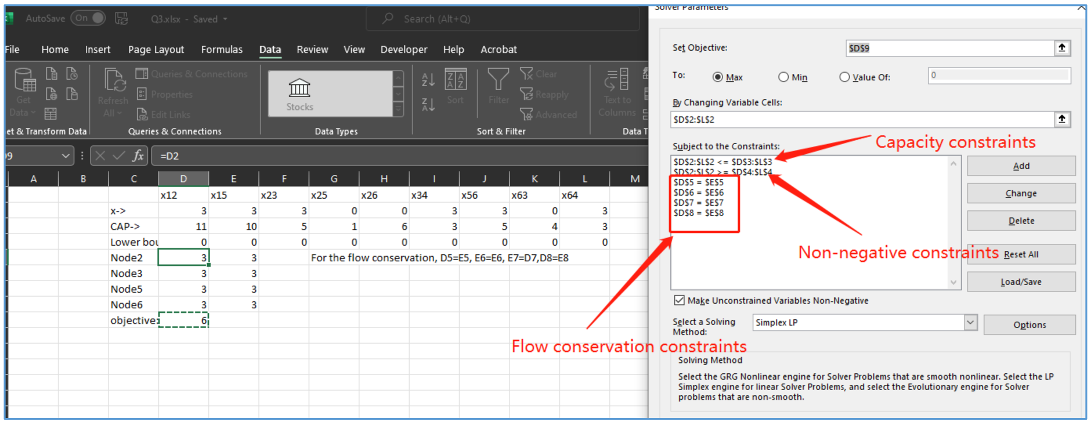
(3) Non-negativity constraints.

$$x_{ij} \geq 0, \forall i, j$$

3.3. (7 marks) Solve this optmization problem using Excel Solver, and provide the optimal values of the objective function and the decision variables.

Remark: *Name the Excel file "*Q3_Max_Flow*" and upload it with your answers on Canvas.*

See the attached excel file named Q3_Max_Flow.xlxs. The objective is 6, and the value of x_{ij} is shown in the figure.



3.4. (5 marks) Use the max flow min cut theorem to verify that the solution is correct. List all the cuts of this graph.

There are in total 9 cuts listing as follows, (also with total data rate of each cut).

Cut 1: (1,2), (1,5); total data rate: 21;
Cut 2: (2,3), (2,6), (2,5), (1,5): total data rate: 22;
Cut 3: (1,2), (5,6): total data rate: 16;
Cut 4: (3,4), (2,6), (1,5), (2,5): total data rate: 21;
Cut 5: (2,3), (2,6), (5,6): total data rate:16;
Cut 6: (1,2), (6,3), (6,4): total data rate:18;
Cut 7: (2,3), (6,3), (6,4): total data rate: 12;
Cut 8: (3,4), (5,6), (2,6): 14;
Cut 9: (3,4), (6,4): 6.

The min-cut is Cut 9, with the total data rate of 6, is consistent with the results obtained by the Excel file.

<u>Question 4 (15 marks)</u>

The graph below (Figure 4) represents a network owned by an ISP that connects City 1, City 2, and City 3. VTE corporation has offices in all these three cities.

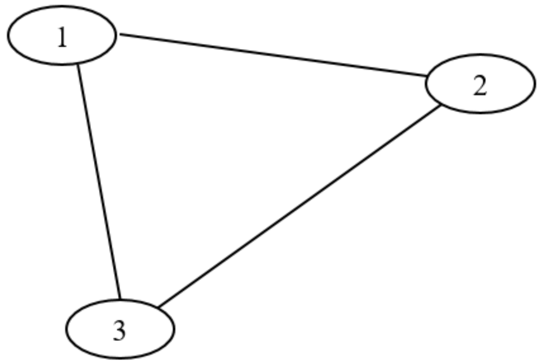


Figure 4

The link that directly connects City *i* and City *j* is called (*i*, *j*).

The traffic demands between the offices in City 1, City 2 and City 3 are given in the following table. The units of the traffic demands are Tb/s (terabit per second).

OD Pair	Traffic Demand
1 2	*N5* Tb/s
2 3	*N6* Tb/s
1 3	*N7* Tb/s

Notice that we consider un-directional traffic demands for simplicity. They represent the total traffic demands in both directions for each origin-destination pair.

Traffic between offices in two cities always uses the shortest path (that minimizes the number of links on the path). That is, traffic between City *i* and City *j* will use Link (*i*, *j*). But if Link (*i*, *j*) fails, it will use a 2-hop path through the third city.

The cost of 1 Tb/s on any link is equal to $30,000 per month.

4.1. (1 mark) Write down the values of *N5*, *N6*, and *N7*.

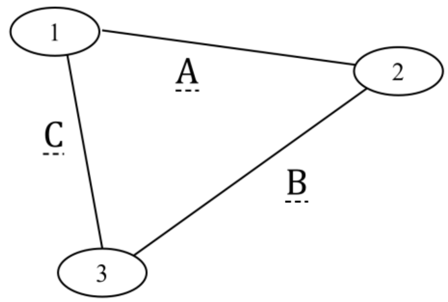
Remarks: *The value Ni is the ith digit of your 10-digit number.*

N5 = 4, *N6* = 3, *N7* = 3.

4.2. (14 marks) VTE wishes to build a private network that is resilient to any single link failure. Specify the capacity that VTE needs to hire for each link such that the total cost per month is minimized and sufficient capacity is available to route the traffic in the shortest path regardless of any single link failure. Show all steps and provide clear justification for your answers.

Notice: *Provide the numerical results of the optimal capacity of each link and the minimal total cost per month.*

We label the three links with A, B, C. Let $T_{i,j}$ represent the capacity of the path between City i and City j.



OD Pair	Traffic Demand
1 2	4 Tb/s
2 3	3 Tb/s
1 3	3 Tb/s

Based on the above table, we got constrains:

$$T_{1,2} \geq 4$$
$$T_{2,3} \geq 3$$
$$T_{1,3} \geq 3$$

If Link A failed, the path between City 1 and City 2 is $\mathrm{C} \rightarrow \mathrm{B}$ (or $\mathrm{B} \rightarrow \mathrm{C}$).
Then we get

Link	A	B	C
Minimal Capacity	NA	3 + 4 = 7	3 + 4 = 7

If Link B failed, the path between City 2 and City 3 is $\mathrm{A} \rightarrow \mathrm{C}$ (or $\mathrm{C} \rightarrow \mathrm{A}$).
Then we get

Link	A	B	C
Minimal Capacity	4 + 3 = 7	NA	3 + 3 = 6

If Link C failed, the path between City 1 and City 3 is $\mathrm{A} \rightarrow \mathrm{B}$ (or $\mathrm{B} \rightarrow \mathrm{A}$).
Then we get

Link	A	B	C
Minimal Capacity	4 + 3 = 7	3 + 3 = 6	NA

To conclude, the optimal capacity of every link should be:

Link	A	B	C
Optimal Capacity	7	7	7

Thus, the minimal cost per month is given by

$$30000 \times (7 + 7 + 7) = 630000 \text{ (dollars)}$$

Question 5 (13 marks)

Consider the secretary problem as follows.

There are N candidates to be interviewed in a sequence for a job. Let N be given by $N = 30 + N7$. Before the interviews, they are considered to be all equally qualified. Assume that after each interview, the interviewers can tell who is the best candidate among those interviewed, and a hiring decision is thus made that cannot be changed afterward. The aim is to maximize the expected probability to choose the best among N candidates.

Remark: *The value N7 is the 7th digit of your 10-digit number.*

5.1. (3 marks) Provide a dynamic programming formulation for this problem, including

a. Optimal value functions;
b. Recurrent relations;
c. Boundary condition(s).

Clearly explain your formulation in your own words.

(1) Define the two optimal value functions:

$v(m)$ ≡After the mth interview, the probability to choose the best among N candidates if the mth candidate is not selected.

$u(m)$ ≡After the mth interview, the probability to choose the best among N candidates under the optimal policy if the mth candidate is the best among the first m candidates.

(2) Recurrent relations (two equations):

$v(m) = \frac{m}{m+1} v(m+1) + \frac{1}{m+1} u(m+1)$	①
$u\,(m) = \max\left[\frac{m}{N},\ v(m)\right]$	②

Explanation to ①:

Given that the mth candidate is not chosen, the $(m+1)$ th candidate will be either not the best among the first m+1 candidates, with probability $\frac{m}{m+1}$, or will be the best among the first m+1 candidates, with probability $\frac{1}{m+1}$. If the $m+1$ candidate is not the best among the first m+1 candidates, s/he is definitely not selected, then the probability of selecting the best candidate among the N candidates is $v(m+1)$. If the $m+1$ th candidate is the best among the first m+1 candidates, then the probability of selecting the best candidate among the N candidates is $u(m+1)$.

Explanation to ②:

Given that the m th candidate is the best among the first m candidates, then s/he is either chosen or not. This is a decision that needs to be made whenever we have the best among the first m candidates. Recall that if the mth is not the best among the first m candidates, s/he is never selected. This probability of choosing the best among the N candidates is equal to $\frac{m}{N}$ if candidate m is chosen and it is equal to $v(m)$ if candidate m is not chosen. By choosing the maximum of the two, we make the optimal decision.

(3) Boundary condition(s):

$$u\,(N) = 1,$$
$$v(N) = 0.$$

Explanation: Given that candidate N is the best among the first N candidates, then s/he is chosen for sure, and then the probability of choosing the best candidate among the N candidates is equal to 1. Therefore, $u\,(N) = 1$. After the Nth interview, if the Nth candidate is not chosen, then the probability of choosing the best candidate among the N candidates is equal to zero. Therefore, $v(N) = 0$.

5.2. (3 marks) Consolidate your answer in sub-question 5.1 to a single optimal value function with a single boundary condition and recurrent relation. Clearly explain all steps in your own words.

If you use sources for sub-question 5.1 and 5.2, quote your sources.

Substituting ② to ①, we obtain:

$$v(m) = \frac{m}{m+1} v(m+1) + \frac{1}{m+1} \max\left[\frac{m+1}{N}, v(m+1)\right]$$

with the boundary condition $v(N)$=0, we obtain:

$$v(N-1) = 0 + \frac{1}{N}(1) = \frac{1}{N}$$

Source: M. J. Beckman, “Dynamic Programming and the secretary problem” Computer. Math. Applic. vol. 19., no. 11, pp. 25-28, 1990.

5.3. (7 marks) Provide your solution in an Excel file called "Q5_secretary" and upload it with your answers on Canvas.

See attached Excel file called "Q5_secretary". It is for 35 candidates. Adapt it to a different number of candidates.

<u>Final Exam Submission</u>

Before the exam end-time, submit the following:

1. All the pages with your answers to questions 1-6 in a PDF file. The PDF file should also include the declaration "I pledge to follow the Rules on Academic Honesty and understand that violations may lead to severe penalties…." with signature and date.

2. The three Excel files: "Q2_Regression", "Q3_Max_Flow ", and "Q5_Secretary".

3. Submit all these files (the PDF and the three Excel files) to the Canvas Assignment called “Final Exam” separately, **NOT** in a Zip file.

- END -