# THE UNIVERSITY OF AUCKLAND

**SEMESTER ONE 2019**
**Campus: City**

**ENGINEERING SCIENCE**

**Algorithms for Optimisation**

**(Time allowed: THREE hours)**

**NOTE:** The marks for this exam total 180 (being approximately 1 mark per minute).

This exam consists of **THREE** sections worth 60 marks each.

Answer **ALL** questions in **ALL** sections.

Answers for each question should be written on this script in the space provided directly after the question.

If you run out of space for a question, you may request an extra 4-page script book. Any extra booklets must be tied to this exam paper at the conclusion of the exam.

Good luck from Tony, Andrew and Kevin.

**Surname:**________________ **Forename(s):**________________

**University of Auckland ID number:**________________

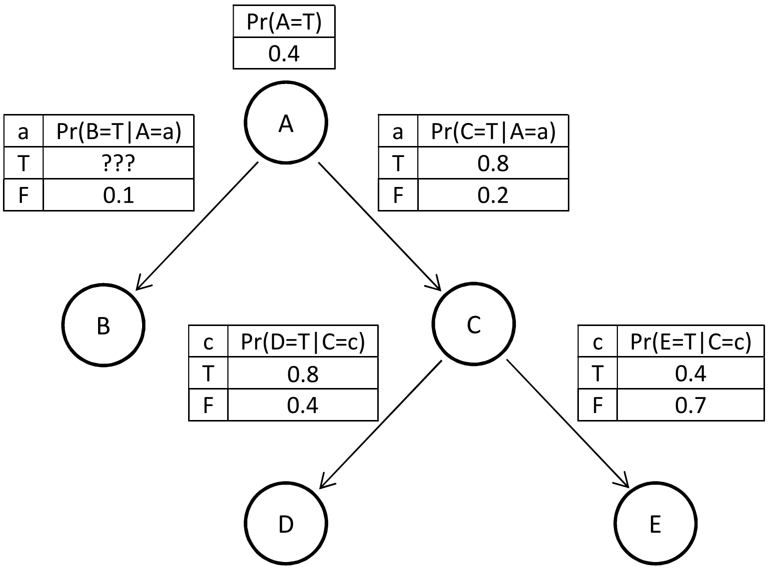
| Section | A | | B | | C | | |
|---|---|---|---|---|---|---|---|
| Question | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Mark | | | | | | | |
| Out of | 18 | 42 | 30 | 30 | 35 | 25 | 180 |

**ID:** ______________________

## SECTION A – Decision Making under Uncertainty (60 marks)

### Question 1 (18 marks)

Consider the Bayesian network given in the diagram below. Each of the random variables {A,...,E} can take a value of either **T** or **F**. The conditional probability of each random variable being **T** is given next to the incoming arc for the corresponding node.




(a) Suppose you are told that, without any further information, the probability that B is **T** is 0.3. What is Pr(B = **T** | A = **T**)? (2 marks)

(b) Suppose we learn that C is **T**; how will this change the probabilities that D and E are each **T**? Explain your answers. (3 marks)

**ID:** ______________________

(c) Suppose we learn that D is **T** (in addition to C being **T**); how will this change the probability that E is **T**? Explain your answer. (2 marks)

(d) Given the information learned in the previous questions, use Bayes' theorem to find the probability that A is **T**. (3 marks)

Now suppose we know that B, D and E are all **T** (but have no direct information about A or C), and we run the belief propagation algorithm, finding that:

$$\lambda(\mathrm{A}) = \begin{bmatrix} 0.1872 \\ 0.0288 \end{bmatrix}, \qquad \pi(\mathrm{C}) = \begin{bmatrix} 0.68 \\ 0.32 \end{bmatrix}$$

Note that **T** corresponds to the first component of the vectors, and **F** the second.

(e) What is the probability that A is **T**? (3 marks)

(f) Use the belief propagation algorithm to find $\lambda(\mathrm{C})$. (3 marks)

**ID:** ____________________

(g) What is the probability that C is **T**? (2 marks)

**ID:** ______________________

## Question 2 (42 marks)

A large manufacturer of consumer electronic products is launching a new type of device and there's a lot of uncertainty about how many they can sell. Each device costs \$150 to make, and they are considering selling these devices for either \$300, \$400 or \$450. They have hired an analytics consultancy, who has provided predictions for the number of these devices that they will be able to sell, given both the price of the device, and the perceived interest on social media platforms. These predictions are that if there is `high` interest then the demand will be given by the equation $d(p) = 25000 - 20p$, whereas if there is `low` interest the demand will be given by the equation $d(p) = 27000 - 40p$, where $p$ is the price the manufacturer sets for the product.

(a) Complete the table, below, showing the profit (in thousands of dollars) that the manufacturer will make, for each decision and state of nature. (5 marks)

| | | Interest | |
|---|---|---|---|
| | | | |
| | \$300 | | 2250 |
| Price | \$400 | 4250 | |
| | \$450 | | 2700 |

(b) Suppose that this manufacturer is pessimistic. Which price (of the three available) would they sell their device for? Explain your answer. (2 marks)

(c) Suppose that this manufacturer wishes to minimise maximum regret. Which price (of the three available) would they sell their device for? Explain your answer. (2 marks)

**ID:** ______________________

In order to better predict the demand, the analytics consultancy is tracking interest in the launch of the product on social media, and they anticipate that by the time the product launches there is a 0.6 probability of `high` interest and a 0.4 probability of `low` interest.

(d) Draw and fully label a decision tree for this problem. Indicate the option that maximises EMV and state the expected profit. (10 marks)

The analytics consultancy states that it can perform additional *machine learning* to give better estimates of the interest in the device. The machine learning model can return either a `positive` or `negative` result; if it returns `positive` then there is a 75% that the interest will be high (and 25% chance it will be low), whereas if it returns `negative` there is a 75% chance that the interest will be low (and 25% chance it will be high).

(e) In order for all the probabilities to be consistent, what is the probability that the machine learning model will return `positive`? (2 marks)

**ID:** ____________________

(f) What is the maximum that the manufacturer should be willing to pay for this additional information? Explain your answer. (4 marks)

Suppose the manufacturer's utility function is $U(\pi) = 1 - e^{-\frac{\pi}{2000}}$, where $\pi$ is their profit (in thousands of dollars), incorporating fixed costs of \$500,000 associated with setting up the production line, and marketing.

(g) Given the utility function and fixed costs above, what is the certainty equivalent value corresponding to the decision to price the device at \$400? (5 marks)

**ID:** ____________________

Suppose that interest in this new type of device changes over time, and can be approximated by a hidden Markov model. The interest each week can be modelled as a Markov chain, with states `low` and `high`, with the following transition matrix.

$$P = \begin{array}{c c} & \begin{array}{cc} \texttt{low} & \texttt{high} \end{array} \\ \begin{array}{c} \texttt{low} \\ \texttt{high} \end{array} & \left[\begin{array}{cc} 0.9 & 0.1 \\ 0.2 & 0.8 \end{array}\right] \end{array}$$

Furthermore, the probability mass functions for the number of social media posts in a week, given the level of interest, are provided in the table below.

| | Interest | |
|---|---|---|
| Number of Posts | `low` | `high` |
| 0 – 10,000 | 0.6 | 0.1 |
| 10,001 – 20,000 | 0.4 | 0.3 |
| 20,001 – 30,000 | 0.0 | 0.6 |

It was originally estimated that the probability that interest was high in week 1 was 0.8. However, the manufacturer has observed 24,521 social media posts about this device in week 1, 18,241 posts in week 2, and 15,346 posts in week 3.

(h) Explain the difference between the Viterbi and Forward-Backward algorithms. If we want to determine the probability that interest was `high` or `low` in a given week, which algorithm should we use here? (3 marks)

(i) Given the observed number of social media posts each week, what is the probability that interest is `high` in week 1? (2 marks)

**ID:** ______________________

(j) Suppose you are given $\beta_{\texttt{high}}(2) = 0.32$ and $\beta_{\texttt{low}}(2) = 0.39$. Apply the appropriate algorithm to determine the probability that interest in the device was `high` in week 2. (5 marks)

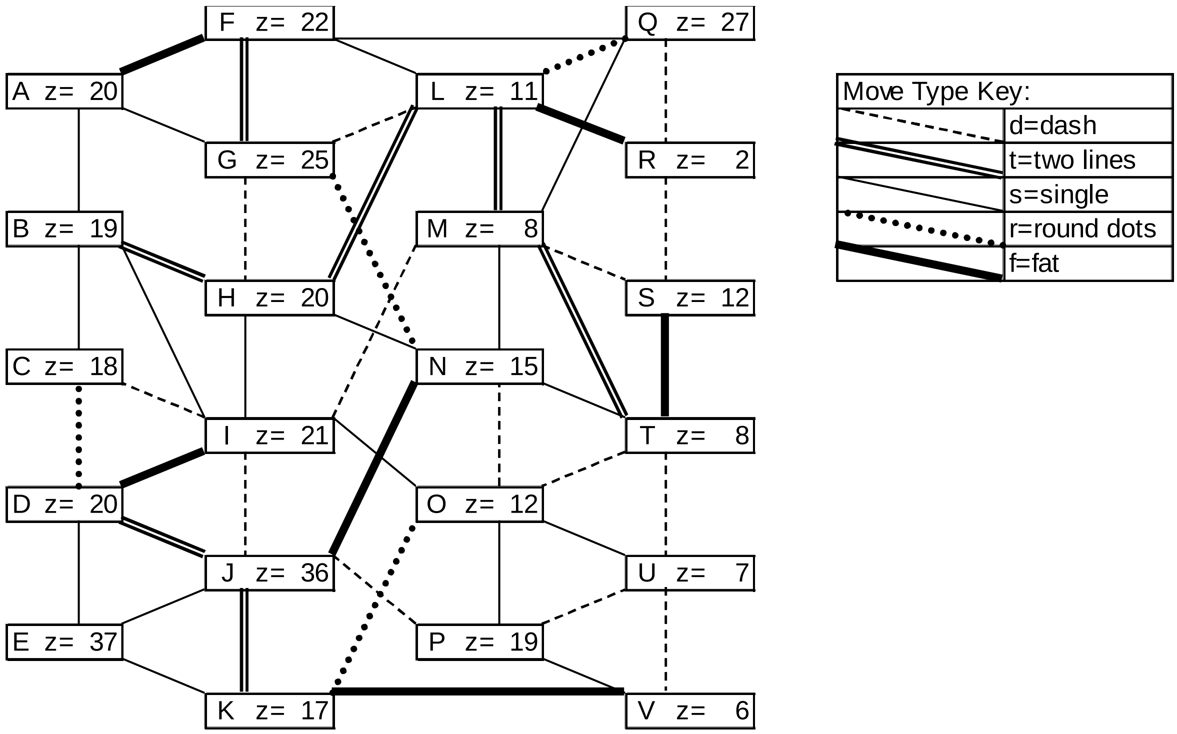
(k) Explain why knowing the number of social media posts in week 3 affects the probability that the interest is `high` in week 2. (2 marks)

**ID:** ______________________

## SECTION B – Heuristics (60 marks)

### Question 3 (30 marks)

The following diagram shows the solution space for a minimisation problem. Each box represents a solution; the solution name (A, B, ..., V) and the solution's objective value z are shown in each box. Where two solutions are joined by a line, they are considered neighbours under some neighbourhood search scheme. The different types of lines indicate the different types of moves that can be made in the local search; these move types are listed in the key. Use this diagram to answer the following questions.




(a) Circle the correct underlined phrase in the following: (2 marks)

Solution R is <u>a local minimum</u> / <u>a global (and local) minimum</u> / <u>neither of these</u>.

Solution C is <u>a local minimum</u> / <u>a global (and local) minimum</u> / <u>neither of these</u>.

Solution N is <u>a local minimum</u> / <u>a global (and local) minimum</u> / <u>neither of these</u>.

(b) What sequence of solutions would be generated using a steepest descent neighbourhood search starting at solution A? (1 mark)

**ID:** ________________________

We now wish to use Tabu Search to solve this problem. Assume each tabu search move is characterised by its type, either dash (d), two lines (t), single (s), round dots (r) or fat (f), and that our tabu list stores the move type for the last two moves (e.g. (r,s) if the last move was s, and the one before that was an r). (If all moves are tabu, a random move is selected.)

(c) Complete the table below to show the sequence of the next 5 solutions that would be visited after the current solution V (giving 6 solutions in total). Be sure to identify both each solution and the contents of the tabu list after each move. You should assume the tabu list is initially empty. (4 marks)

Solution 1: __V__, Tabu list = ()

Solution 2: _____, Tabu list = (_____________)

Solution 3: _____, Tabu list = (_____________)

Solution 4: _____, Tabu list = (_____________)

Solution 5: _____, Tabu list = (_____________)

Solution 6: _____, Tabu list = (_____________)

(d) What is the name of the behaviour that you would observe if you continued this Tabu run? (1 mark)

(e) A friend is going to choose a new Tabu list size of either 1, 3, 4 or 5. Of these, which (if any) would give the *same* search behaviour as you observed above. (2 marks)

We now wish to solve this problem using Simulated Annealing. Recall that Simulated Annealing uses the formula $e^{-[f(y)-f(x)]/T}$.

(f) Suggest an appropriate method from the literature for choosing a starting value for $T$. (2 marks)

**ID:** ____________________

(g) In our simulated annealing run for the problem above, we:

(i) start the search at some random solution, solution R, and then
(ii) move to solution L,
(iii) move to solution M, and then
(iv) reject moving to solution S.

We are using an inhomogeneous cooling schedule that starts with temperature $T = 100$ during the first iteration and then reduces the temperature by 10% at the end of each iteration. Explain how the steps of the simulated annealing algorithm produce this particular sequence of actions. Show the current solution and values of $T$, $f(x)$ and $f(y)$ in the spaces below. (5 marks)

Current solution = __R__, $T =$ __100__

Move to L: $f(x) =$ ________, $f(y) =$ ________

Current solution = _____, $T =$ _____

Move to M: $f(x) =$ ________, $f(y) =$ ________

Current solution = _____, $T =$ _____

Reject moving to S: $f(x) =$ ________, $f(y) =$ ________

Current solution = _____

**ID:** ________________________

(h) Explain why an integer programming solver might use local branching, and give an example of how it might be used given some current solution $x_1 = x_2 = x_4 = 1$, $x_3 = x_5 = 0$ to a binary integer programming problem. (3 marks)

(i) What is RINS (relaxation induced neighbourhood search) used for? Explain how RINS works, detailing the role subMIPs play in RINS and showing how a subMIP is created, using as an example some current fractional solution $x_1 = 0.5, x_2 = 0.5, x_3 = 1, x_4 = x_5 = 0$ and some incumbent best solution $x_1 = 0, x_2 = 1, x_3 = x_4 = x_5 = 0$. (4 marks)

(j) When researchers compare heuristic algorithms, they often do so by comparing the final solutions they obtain for a set of test problems using each of the different heuristics being tested. What is wrong with this approach? Give a better comparison technique. (2 marks)

**ID:** ____________________

(k) Explain what is meant by 'diversification' and 'intensification' in the context of heuristic search. (2 marks)

(l) Describe how the degree of diversification and the degree of intensification change during a Simulated Annealing run, and name the parameter that Simulated Annealing uses to cause these changes. (2 marks)

**ID:** ______________________

## Question 4 (30 marks)

Consider the sample figure shown below that represents shares being transferred over time. We wish to use local search to change the vertical positions of the nodes (termed 'vertices') to create a new figure where the lines (termed 'arcs') between vertices are as near to flat (i.e. near to horizontal) as we can achieve. Note that moving a vertex up or down also moves the ends of the arcs that connect to that vertex.

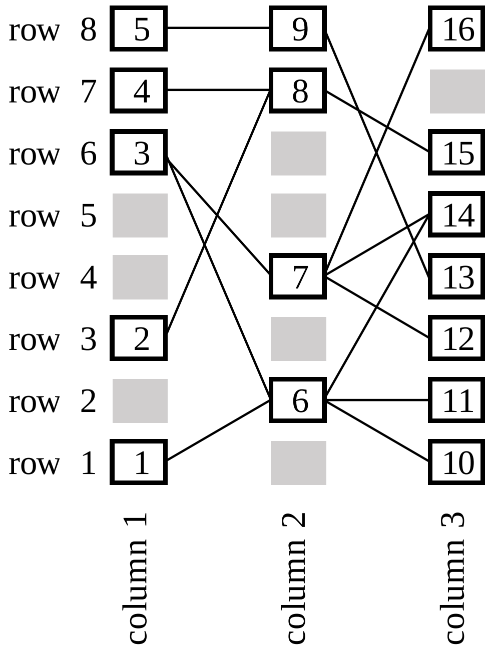
This optimisation problem is defined using the following notation. We have:

- $C$ columns numbered $1, 2, ..., C$,
- $R$ rows numbered $1, 2, ..., R$, and
- a set $V = \{1, 2, ..., n\}$ of $n$ vertices

Each vertex $v \in V$ has:

- an associated column $c_v \in \{1, 2, ..., C\}$
- a (possibly empty) set of predecessor vertices $P_v \subseteq V$
- a (possibly empty) set of successsor vertices $S_v \subseteq V$

Vertex $v$ in column $c_v$ is connected by 'predecessor arcs' to each predecessor vertex $u \in P_v$ in the previous column $c_u = c_v - 1$, and is connected by 'successor arcs' to vertices $u \in S_v$ in the next column $c_u = c_v + 1$.




In the example shown here, we have $C = 3$ columns, $R = 8$ rows, and $n = 16$ vertices. Vertex 8, for example, has associated column $c_8 = 2$, predecessor vertices $P_8 = \{2, 4\}$ (defining the predecessor arcs 2-8 and 4-8) and successor vertices $S_8 = \{15\}$ (defining successor arc 8-15). Vertex 3 has $c_3 = 1$, $P_3 = \{\}$, and $S_3 = \{6, 7\}$.

We represent a solution in the form $\mathbf{x} = (x_1, x_2, ..., x_n)$, where $x_v \in \{1, 2, ..., R\}$ is the row assigned to vertex $v$. Two vertices in the same column cannot be in the same row.

For any vertex $v \in V$, we define $f_v(\mathbf{x}) = \sum_{u \in P_v} |x_u - x_v|$ as a measure of the flatness of vertex $v$'s predecessor arcs.

(a) Using the notation above, give a general formula for the total flatness $f(\mathbf{x})$ of all the arcs in a problem for some solution $\mathbf{x}$. (2 marks)

$f(\mathbf{x}) =$

(b) Consider two vertices $u$ and $v$ in the same column, i.e. where $c_u = c_v$. We wish to define a neighbouring solution in which the rows of any two such vertices are swapped. Using the notation defined above, give a formal definition of this neighbouring solution $\mathbf{y} = (y_1, y_2, ..., y_n) = \mathbf{y}(\mathbf{x}, ...)$. Be sure to replace the '...' with appropriate arguments, and to give a formal definition for the elements $y_i$ of $\mathbf{y}$ in terms of these arguments. (2 marks)

**ID:** ____________________

$\mathbf{y}(\mathbf{x}, .......) =$

(c) By using your $\mathbf{y}(\mathbf{x}, ...)$ from above, give a formal definition of the 'swap' neighbourhood $N(\mathbf{x})$ formed from all possible swaps of the type in question (b). Give ranges for your arguments used in $\mathbf{y}(\mathbf{x}, ...)$, and specify any conditions the arguments must satisfy, to ensure each neighbouring solution is included exactly once, and invalid neighbours are excluded. (3 marks)

$N(\mathbf{x}) =$

(d) Give a formula for the change in objective function, $\Delta_v(\mathbf{x}, r)$, if some vertex $v$ is moved to some new row $r$ in the same column. (Assume the row in this column is empty, i.e. this possible vertex position has no other vertex assigned to it.) (3 marks)

$\Delta_v(\mathbf{x}, r) =$

(e) Give a formula that uses the $\Delta_v()$, $v = 1, 2, ..., n$ functions to calculate the change in objective function $\Delta f(\mathbf{x}, u, v)$ if two vertices $u$ and $v$ in the same column have their rows swapped. (2 marks)

$\Delta f(\mathbf{x}, u, v) =$

(f) We wish to extend our neighbourhood to allow vertices to be moved to an empty row in their column. A friend has suggested that the easiest way to do this is to add dummy vertices to the problem. Explain exactly how this would be done, and what the predecessor and successor sets would be for these dummy vertices. (3 marks)

**ID:** ______________________

(g) We wish to use an alternative neighbourhood $N'(\mathbf{x})$ in which every rearrangement of three vertices (including simple pairwise swaps) is considered for every column. We want to better understand the sets of local optima $L$ and $L'$ that would be found using neighbourhood rules $N()$ and $N'()$, respectively. Of the following, circle all that are true in general. (2 marks)

$L \subseteq L'$, $L' \subseteq L$, $L = L'$

(h) Briefly explain how a 'Large Neighbourhood Search' (LNS) operates in general, and then outline problem-specific procedures for the two key LNS operations that could be used to apply the LNS algorithm to this problem. (3 marks)

(i) Briefly detail an appropriate construction heuristic that would be suitable to use as part of a GRASP implementation for this problem. (3 marks)

(j) What is the key advantage of GRASP over repeated random restarts? (1 mark)

**ID:** ____________________

We wish to use a Genetic Algorithm to solve this problem. A friend Sue has suggested that the 'building blocks' for this problem are the vertex assignments in each pair of adjacent columns (i.e. in (column 1, column 2), (column 2, column 3), and so on) as these define the flatness of the arcs between the two columns.

(k) Explain briefly what Sue means by 'building blocks', and why thinking about building blocks is important for designing a crossover operator. (2 marks)

(l) Consider some problem with $R = 5$ rows and $C = 6$ columns for which we have two solutions $\mathbf{p}$ and $\mathbf{q}$ as given below. Note that in these vectors, the vertices are ordered by increasing column $c_v$, with ';' denoting the change from one column to the next. Using words, outline a crossover operator for this problem motivated by Sue's observation, and show one child that might be produced by this operator for parents $\mathbf{p}$ and $\mathbf{q}$. (3 marks)

Operator description:

Example crossover:

$$\mathbf{p} = (\overbrace{1,2,3,4,5}^{\textbf{column 1}};\overbrace{1,2,3,4,5}^{\textbf{column 2}};\overbrace{1,2,3,4,5}^{\textbf{column 3}};\overbrace{1,2,3,4,5}^{\textbf{column 4}};\overbrace{1,2,3,4,5}^{\textbf{column 5}};\overbrace{1,2,3,4,5}^{\textbf{column 6}})$$

$$\mathbf{q} = (4,5,1,3,2;4,3,2,1,5;5,2,1,3,4;1,3,5,4,2;2,5,1,4,3;5,1,2,4,3)$$

$$\text{child} =$$

(m) Explain what difficulty would arise if a standard 2-point crossover operator was applied to parents $\mathbf{p}$ and $\mathbf{q}$. (1 mark)

**ID:** ______________________

## SECTION C – Dynamic Programming (60 marks)

### Question 5 (35 marks)

Chartreuse Education is piloting a 24/7 online tutoring service targeted at students worldwide. A trained tutor can start tutoring on the platform by paying Chartreuse a login charge of \$25, and can tutor as long as they like until they log off (which costs nothing to the tutor). A tutor can choose to login at the start of each hour, if they are not already logged in, and log in as many times as they like during the day, as long as they pay the charge each time.

Jamie, a tutor on the platform, would like to plan his next 24 hours of tutoring. He will make a decision at the start of each hour, of whether he chooses to log into the tutoring platform that hour. Chartreuse Education requires that a tutor must be actively tutoring for at least 10 minutes in each hour that they are logged into the platform, and, for health and safety reasons, they must take a break of at least 5 minutes in each hour they are logged in. Based on his knowledge of the private tutoring market and worldwide income distribution, Jamie has calculated the price that he should expect to receive per minute of tutoring in each of the next 24 hours, $p(t)$, where $t = 1, 2, ..., 24$. *While actively tutoring,* Jamie values his time at $c = \$0.20$ per minute regardless of the time of day.

Jamie would like to schedule his tutoring to maximise the amount of profit made (revenue minus costs, including the value of Jamie's time) over this 24 hour period. You may assume that Jamie is not tutoring immediately before the start of the 24 hour period, and that he must log off the system at the end of the period. Jamie is not concerned about the profit in the future days. Jamie would like to formulate this problem as a dynamic program.

(a) What are the stages of this problem? (2 marks)

(b) Suppose Jamie is tutoring this hour, and that $p(t) > c$ for that hour. How much time should Jamie spend tutoring that hour to maximise his profit? (2 marks)

(c) Suppose that Jamie is tutoring in hour $t$. Write down an expression for the optimal profit in this hour, $\Pi(t)$, and briefly explain this result.
You may find the notation $(x)_+ = \max\{x, 0\}$ helpful in your answer. (4 marks)

**ID:** ______________________

(d) Let $z = 1$ if Jamie is tutoring, and 0 if he's not.
Let $V_t(z)$ denote the maximum profit that can be earned in hours $t, t+1, ..., 24$ given that Jamie is in state $z$ at the beginning of hour $t$.

(i) What is $V_{25}(0)$? (2 marks)

(ii) What does $V_{25}(1)$ represent in the context of this problem?
What is its value? (3 marks)

(iii) Write down the recursion that $V_t(z)$ must satisfy. You may use $\Pi(t)$, as defined in part (c), in your answer.
Hint: you should have an expression for each state of the system. (6 marks)

**ID:** ____________________

Jamie has also been recruited by Chartreuse Education to screen potential candidates for tutoring roles. There are currently 10 vacancies for tutors that need to be filled. A tutor application is expected to arrive once a day for the next 20 days. Chartreuse gives each potential tutor a quality score $q$, which is assumed to be a random observation of the random variable $Q$, which has a Uniform distribution on the interval $[0, 10]$. This quality is independent from tutor-to-tutor. However, a tutor's quality is not known in advance; it is only realised when an application arrives. If a potential tutor is turned away, they will be hired by their competitor Impulse Tutoring and cannot be re-considered. Once all 10 spots are filled, Chartreuse Education cannot consider any further candidates (i.e. each of them must be rejected regardless of their quality score).

Suppose the stages of the problem are $t = 1, 2, ..., 20$, and let $\mathcal{X}$ represent the set of all possible states of this process, which is the number of vacancies remaining to be filled $\{0, 1, ..., 10\}$, and $x \in \mathcal{X}$. Let $V_t(x, q)$ be the expected sum of the quality scores of the tutors that are hired from those applications received in days $t, t+1, ..., 20$, given that day $t$'s application has been received and a quality score $q$ assigned. Further, let $\hat{V}_t(x) = \mathbb{E}_q[V_t(x, q)]$, the expected sum of quality scores of hired tutors from day $t$ to day 20.

*Recall:* the probability density function of a Uniform distribution on the interval $[a, b]$ is $\frac{1}{b-a}$.

(e) Write down the stage and state at which we would like to ultimately evaluate the expected sum of quality scores. (2 marks)

(f) What is $\hat{V}_{21}(x)$ for all $x \in \mathcal{X}$? (2 marks)

(g) What is $V_{20}(x, q)$ for all $x \in \mathcal{X}$? (2 marks)

(h) Write down a recursion for $V_t(x, q)$ in terms of $\hat{V}_{t+1}(x)$ and $\hat{V}_{t+1}(x - 1)$. Do not attempt to solve this. (2 marks)

**ID:** ____________________

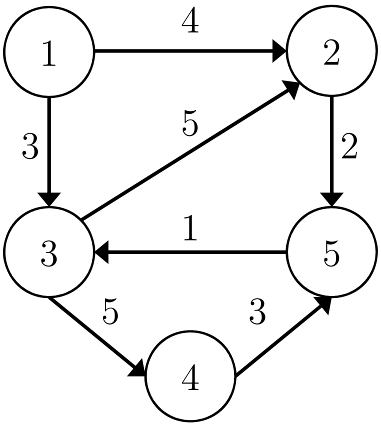
(i) Hence explain briefly why, on day $t$, a tutor will be hired if their quality score is greater than $\pi_t = \hat{V}_{t+1}(x) - \hat{V}_{t+1}(x-1)$, assuming $x \geq 1$. (2 marks)

(j) Therefore, show that $\hat{V}_t(x) = 5 + \dfrac{\pi_t^2}{20} + \hat{V}_{t+1}(x-1)$. (6 marks)

**ID:** ________________________

**Question 6 (25 marks)**

Consider the following directed network with five nodes, with the cost of traversing each arc given as the arc weights:

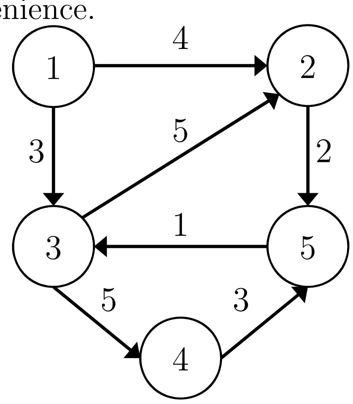



Suppose we seek an infinite path starting at any node that has minimum discounted cost, given a discount factor, $\rho$.

(a) Write down a linear program with seven variables $x_{ij}$ that can be used to solve this problem. (6 marks)

**ID:** ____________________

Suppose the linear program from part (a), with $\rho = 0.2$, was solved and the following output was obtained. A few values are missing. The network diagram is also reproduced for your convenience.




Variable Cells

| Name | Final Value |
|---|---|
| x12 | 0 |
| x13 | 1 |
| x25 | 1.298387 |
| x32 | **** |
| x34 | **** |
| x45 | 1 |
| x53 | 1.459677 |

Constraints

| Name | Shadow Price |
|---|---|
| Node 1 | 4.096774 |
| Node 2 | 2.419355 |
| Node 3 | 5.483871 |
| Node 4 | 3.419355 |
| Node 5 | 2.096774 |

(b) What is the cost of a discounted infinite shortest path starting at node 5? (2 marks)

(c) In the optimal policy, which node is the next node to visit after node 3? Show working to justify your answer. (5 marks)

(d) If we started at Node 1, would we ever re-visit it? Explain your answer briefly. (2 marks)

**ID:** ______________________

(e) Suppose we are starting our discounted infinite shortest path at node 1. How many times will node 4 be visited in this shortest path? Justify your answer. (3 marks)

(f) Explain what *optimal substructure* and *overlapping subproblems* mean in the context of solving a general problem using dynamic programming. Compare and contrast this against the concept of *divide-and-conquer*, and hence describe how a problem exhibiting the first two features can be solved efficiently using dynamic programming techniques. (7 marks)

**ID:** ____________________

**THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK.**

**IT WILL NOT BE MARKED.**