Model solution – Q1

1. (a)
   - Possible advantages over direct methods, e.g. LU decomposition: (i) reduced memory usage, (ii) faster (reduced algorithmic complexity), (iii) some can be easily parallelised, (iv) can terminate early if high accuracy is not needed, (v) can often solve singular problems (with known nullspace).
   - Possible disadvantages over direct methods: (i) slow to converge or may fail for some matrices, e.g. poorly conditioned matrices, (ii) maybe applicable only to specific matrix type, e.g. conjugate gradient method is only for symmetric positive-definite matrices, (iii) can be difficult to predict/model performance *a priori*.

   (b) i. The columns of $\boldsymbol{A}$ sum to one, hence the rows of $\boldsymbol{A}^T$ will sum to one, hence:

$$\boldsymbol{A}^T \mathbf{1} = \mathbf{1}. \tag{1}$$

This implies that $\lambda = 1$ is an eigenvalue of $\boldsymbol{A}^T$. Since the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{A}^T$ are the same, $\lambda = 1$ must be an eigenvalue of $\boldsymbol{A}$.

ii. Since the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{A}^T$ are the same, setting $\boldsymbol{B} = \boldsymbol{A}^T$, for an eigenvector $\boldsymbol{y}$ of $\boldsymbol{B}$, we have

$$\boldsymbol{B}\boldsymbol{y} = \lambda \boldsymbol{y},$$

which can be expressed in terms of components as $\lambda y_i = \sum_j b_{ij} y_j$. It then follows that

$$\begin{aligned} |\lambda y_i| = |\lambda||y_i| &= |\sum_j b_{ij} y_j| \\ &\leq \sum_j b_{ij} |y_j| \\ &\leq |\boldsymbol{y}|_\infty \end{aligned}$$

for all i (recall that $|\boldsymbol{y}|_\infty = \max_i |y_i|$). The last inequality follows from $b_{ij} \geq 0$ and $\sum_j b_{ij} = 1 \ \forall i$. The above inequality can be true only if $|\lambda| \leq 1$

iii. If a vector is repeatedly multiplied by a matrix, the resulting vector will tend towards the eigenvector associated with the largest absolute eigenvector. This is known as *power iteration*:

$$\boldsymbol{x}_{n+1} = \boldsymbol{A}^n \boldsymbol{x}_0$$

or equivalently (in the case of exact arithmetic)

$$\boldsymbol{x}_{n+1} = \frac{\boldsymbol{A}\boldsymbol{x}_n}{|\boldsymbol{A}\boldsymbol{x}_n|}$$

where $\boldsymbol{x}_{n+1}$ is an approximation of an eigenvector of $\boldsymbol{A}$.
For the matrix

$$\boldsymbol{A} = \begin{bmatrix} 0.0 & 0.3 \\ 1.0 & 0.7 \end{bmatrix} \tag{2}$$

we have

$$\boldsymbol{A}^3 = \begin{bmatrix} 0.21 & 0.237 \\ 0.79 & 0.763 \end{bmatrix}$$

and

$$\boldsymbol{A}^3 \begin{bmatrix} 1 & 1 \end{bmatrix}^T = \begin{bmatrix} 0.447 & 1.553 \end{bmatrix}^T = \boldsymbol{x}_3 \approx \boldsymbol{x}$$

The actual eigenvector is in the direction $\begin{bmatrix} 0.3 & 1 \end{bmatrix}^T$. Scaled similarly this approximation is $\begin{bmatrix} 0.288 & 1 \end{bmatrix}^T$, so not a bad approximation.

Assessor's Comments:

61 attempts, Average mark 10.0/20, Maximum 18, Minimum 1.

The low average on this question reflects the high number of partial attempts.

Few answers to part (a) mentioned most of the salient points expected.

Several (not very successful) answers were based more on the content of the Optimization part of 3M1 than the Linear Algebra part. Discussions of the effectiveness of Newton's Method are not relevant to the solution of $\boldsymbol{Ax} = \boldsymbol{b}$…

Few answers to part (b)(i) used the hint in the expected way. Most sought (successfully) to show that the matrix $\boldsymbol{A} - \boldsymbol{I}$ was singular and thus that 1 was an eigenvalue.

Several answers to part (b)(ii) involved circular arguments, often in the context of the $\boldsymbol{l}_2$ norm.

A number of answers to (b)(i) and (b)(ii) took $\boldsymbol{A}$ to be a 2×2 and then made dubious claims about the generalisability of the proof given.

Approaches in part (b)(iii) were often not properly justified or not checked against the (easily found) eigenvector corresponding to an eigenvalue of 1 to confirm that a good approximation was indeed being found.

Some candidates tried to apply Richardson's method, which solves $\boldsymbol{Ax} = \boldsymbol{b}$ rather than $\boldsymbol{Ax} = \boldsymbol{x}$, in part (b)(iii) rather than the expected power iteration method.

Q2

(a) The penalty function method (PFM) penalizes the objective function if the solution lies outside the feasible region.

The barrier function method (BFM) attempts to restrict the search to the feasible region by penalizing solutions that approach its edge.

The PFM can handle both equality and inequality constraints.

The BFM can only handle inequality constraints.

The BFM requires a feasible starting point; the PFM does not.

The PFM can iterate through infeasible space; the BFM cannot.

The BFM always yields a feasible answer; the PFM may not. [20%]

(b) Minimize $f(r_o) = \rho\pi\left(r_o^{\,2} - r_i^{\,2}\right)$

subject to $g_1(r_o) = 2.0 \times 10^3 - U(r_o) \le 0$

where $U(r_o) = \dfrac{1}{r_i}\left[\dfrac{1}{r_i h_i} + \dfrac{1}{r_o h_o}\right]^{-1}$

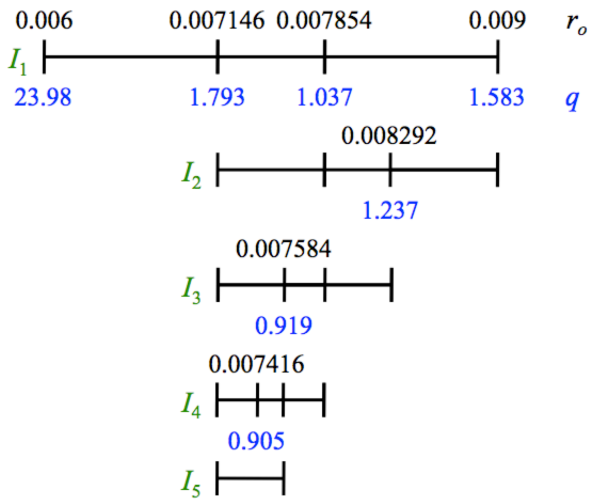
and $g_2(r_o) = r_i - r_o < 0$ [10%]

(c) Minimize $q(r_o, p_k) = f(r_o) + p_k P(r_o)$

where $P(r_o) = \max\left[0, g_1(r_o)\right]^2$

The bound g_2 can be satisfied by restricting the search space appropriately and does not require an additional penalty term. [10%]

(d) From the 3M1 data sheet for the Golden Section method $\dfrac{\Delta x}{d} = 0.382$. Thus in this case the search pattern is:



$\therefore \quad I_5 = 7.146 \le r_o \le 7.584$ mm [40%]

r_o	$f(r_o)$	$U(r_o)$	$p_1 P(r_o)$	$q(r_o, p_1)$
0.006	0.311017667	1846.153846	23.66863905	23.97965672
0.009	1.58336267	2117.647059	0	1.58336267
0.007146	0.736979435	1967.511013	1.055534262	1.792513697
0.007854	1.037252855	2030.506722	0	1.037252855
0.008292	1.237207471	2065.769806	0	1.237207471
0.007584	0.919398083	2007.411329	0	0.919398083
0.007416	0.848146762	1992.477163	0.056593079	0.904739842

(e) The table above (used in the GS search) shows that the constraint will be active at the optimum (the objective function changes monotonically with r_o). In this case the constraint equation can be solved to find the optimal value of r_o.

$$\frac{1}{r_i}\left[\frac{1}{r_i h_i}+\frac{1}{r_o h_o}\right]^{-1}=2000$$

$$\therefore \quad \frac{1}{r_o}=\frac{h_o}{r_i}\left[\frac{1}{2000}-\frac{1}{h_i}\right]=\frac{4000}{0.005}\left[\frac{1}{2000}-\frac{1}{3000}\right]$$

$$\therefore \quad r_o=0.0075 \text{ m or } 7.5 \text{ mm}$$ [20%]

<u>*Assessor's Comments:*</u>

45 attempts, Average mark 10.7/20, Maximum 20, Minimum 1.

The least popular question, only attempted by about half the candidates. There were some very good attempts, but a lot of partial answers.

A litany of avoidable errors featured in answers.

Few answers to (a) covered all the salient points.

Some candidates implemented a barrier function rather than a penalty function, even though formulations for both are given on the 3M1 Data Sheet. Others failed to implement the penalty function correctly despite the standard formulation being given. Several candidates failed to square the term quantifying the extent of constraint violation.

Many candidates reformulated the objective function and constraint, which was not required. Although they were given full credit if they still found the correct solution/interval, in all too many cases the reformulations were incorrect, resulting in the wrong problem being solved.

Several candidates failed to recognise that their formulation did not result in a function with a minimum within the initial interval. Others did not implement the penalisation term correctly, penalizing solutions that satisfied the constraint as well as those that violated it.

Poor attention to detail and poor presentation of Golden Section line search calculations made it hard to allocate partial credit to some incorrect solutions as it was not clear where things had gone wrong.

Some answers to part (d) converged to an infeasible solution because the penalisation was too light for the candidate's reformulated problem. This was not penalised in the mark scheme if the candidate recognised this and commented appropriately.

In part (e) the correct solution was sometimes given but without adequate justification for full credit. Other answers implied no realisation that the problem was univariate – discussing the use of multivariate search methods. Others failed to consider the question of feasibility.

Q3

(a) Total production costs in the three months are:

$$(x^2+1000)+(y^2+1000)+(z^2+1000)$$

Noting that there is no hiring/firing at the end of month three, the hiring/firing costs are:

$$10(x-y)^2+10(y-z)^2$$

The storage costs for any excess production after each month are:

$$20(x-40)+20(x+y-100)+20(x+y+z-150)$$

Thus the total costs (to be minimized) are:

$$f(x,y,z)=(x^2+1000)+(y^2+1000)+(z^2+1000)+10(x-y)^2+10(y-z)^2$$
$$+20(x-40)+20(x+y-100)+20(x+y+z-150)$$

$$\therefore \quad f(x,y,z)=x^2+y^2+z^2+10(x-y)^2+10(y-z)^2+60x+40y+20z-2800$$

Thus
$$C=2800$$

The contract to supply represents the constraints to be satisfied – total production must exceed the number of units needed to fulfil the contract at the end of each month. Thus the minimization is subject to:

$$g_1(x,y,z)=40-x\le 0 \qquad \text{(G1)}$$

$$g_2(x,y,z)=100-x-y\le 0 \qquad \text{(G2)}$$

$$g_3(x,y,z)=150-x-y-z\le 0 \qquad \text{(G3)}$$

[20%]

(b) Using Kuhn-Tucker multipliers the appropriate 'Lagrangian' is:

$$L=f+\mu_1 g_1+\mu_2 g_2+\mu_3 g_3$$

and the Kuhn-Tucker optimality conditions are:

$$\frac{\partial L}{\partial x}=2x+20(x-y)+60-\mu_1-\mu_2-\mu_3=0$$

$$\therefore \quad \frac{\partial L}{\partial x}=22x-20y+60-\mu_1-\mu_2-\mu_3=0 \qquad (3.1)$$

$$\frac{\partial L}{\partial y}=2y+20(y-x)+20(y-z)+40-\mu_2-\mu_3=0$$

$$\therefore \quad \frac{\partial L}{\partial y}=42y-20x-20z+40-\mu_2-\mu_3=0 \qquad (3.2)$$

$$\frac{\partial L}{\partial z}=2z+20(z-y)+20-\mu_3=0$$

$$\therefore \quad \frac{\partial L}{\partial z}=22z-20y+20-\mu_3=0 \qquad (3.3)$$

$$\mu_1 g_1=\mu_1(40-x)=0 \qquad (3.4)$$

$$\mu_2 g_2=\mu_2(100-x-y)=0 \qquad (3.5)$$

$$\mu_3 g_3=\mu_3(150-x-y-z)=0 \qquad (3.6)$$

[20%]

(c) If $x = 40$, $y = 60$ and $z = 50$, then equations (3.4) to (3.6) are satisfied. Equation (3.3) gives

$$\mu_3 = 22z - 20y + 20 = 22 \times 50 - 20 \times 60 + 20 = -80$$

As the values of all the μ_i must be ≥ 0 at a minimum, this solution cannot represent the optimum. [10%]

(d) Given that at the optimum $\mu_1 = 0$ and $\mu_2 > 0$, there are two possible cases to be investigated:

$\mu_1 = 0,\ \mu_2 > 0,\ \mu_3 = 0$ and $\mu_1 = 0,\ \mu_2 > 0,\ \mu_3 > 0$

For $\mu_1 = 0,\ \mu_2 > 0,\ \mu_3 = 0$:

Equation (3.1) gives $\quad 22x - 20y + 60 = \mu_2$

Equation (3.2) gives $\quad 42y - 20x - 20z + 40 = \mu_2$

Equation (3.3) gives $\quad 22z - 20y + 20 = 0$

And equation (3.5) gives $\quad x + y = 100$

These simultaneous equations can be solved using the appropriate functionality on the CUED-approved calculator. The solution is

$$x = \tfrac{12045}{236} = 51.038$$

$$y = \tfrac{11555}{236} = 48.962$$

$$z = \tfrac{56595}{1298} = 43.602$$

$$\mu_2 = \tfrac{24025}{118} = 203.602$$

At first glance this appears to be a viable optimum as $\mu_2 > 0$ but careful inspection shows that $x + y + z < 150$, so constraint equation (G3) is violated for this solution. [20%]

For $\mu_1 = 0,\ \mu_2 > 0,\ \mu_3 > 0$:

Equation (3.1) gives $\quad 22x - 20y + 60 = \mu_2 + \mu_3$

Equation (3.2) gives $\quad 42y - 20x - 20z + 40 = \mu_2 + \mu_3$

Equation (3.3) gives $\quad 22z - 20y + 20 = \mu_3$

Equation (3.5) gives $\quad x + y = 100$

And equation (3.6) gives $\quad x + y + z = 150$

Again, these simultaneous equations can be solved using the appropriate functionality on the CUED-approved calculator. The solution is

$$x = \tfrac{1295}{26} = 49.808$$

$$y = \tfrac{1305}{26} = 50.192$$

$$z = 50$$

$$\mu_2 = \tfrac{465}{13} = 35.769$$

$$\mu_3 = \tfrac{1510}{13} = 116.154$$

This is a possible optimum: $\mu_2 > 0$, $\mu_3 > 0$ and the three constraint equations are all satisfied. [20%]

For this solution, given that x, y and z must in practice be integers $x = y = z = 50$ is the production schedule to implement. The values of μ_2 and μ_3 indicate that the objective function could be improved more readily by relaxing constraint g_3 rather than g_2. This is because for the problem as posed there is no firing/hiring penalty associated with the end of month three. [10%]

Assessor's Comments:

80 attempts, Average mark 13.3/20, Maximum 20, Minimum 4.
A popular question attempted by 88% of candidates.

The only consistent failing in part (a) was that the costs associated with storage of excess units were not always convincingly explained.

Several candidates failed to mention the requirement that $\mu_i g_i = 0$ for all inequality constraints in their answers to part (b).

Some candidates incorrectly thought that an inequality constraint being active meant the corresponding KT multiplier was zero.

Others thought that $\nabla f = 0$ was a sufficient condition for a minimum even though the problem was clearly constrained.

Several attempts were disadvantaged by incorrectly formulated K-T optimality conditions despite the general form being given on the 3M1 Data Sheet. Solutions that were otherwise correct were not heavily penalised unless the resulting 'optimal' solution was clearly not optimal (e.g. due to constraints being violated).

Many candidates incorrectly thought that second-order optimality conditions needed to be checked, failing to recall that for a problem with no equality constraints the first-order KT optimality conditions are sufficient.

Most candidates did not take advantage of the simultaneous equation solving functionality on their CUED-approved calculator. Many attempts then floundered due to sloppy algebra.

Several candidates only tested one of the two possible solution paths in part (d): $\mu_3 = 0$ or $\mu_3 > 0$.

Question 4 *Discrete State Space Models*

(a) Every element of $\mathbf{P}$ must be non-negative and the sum of the elements in each row must be 1. [10%]

(b) The general expression to be satisfied is

$$\pi = \pi\mathbf{P}$$

However, as the structure of the matrix means that the first m states are *transient*, then the expression can be simplified to

$$\pi_1 = \mathbf{0}; \quad \pi_2 = \pi_2\mathbf{P}_{22}$$

The distribution does not depend on the initial state if the states associated with $\mathbf{P}_{22}$ are recurrent. [20%]

(c) If $\mathbf{P}_{12} = \mathbf{0}$ then there may be two distinct stationary distributions, one based on $\mathbf{P}_{11}$ the other on $\mathbf{P}_{22}$. The selection of the stationary distribution depends on the initial state in this case. [15%]

(d) (i) The transition matrix is

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & a & 1-a \\ 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix}$$

[15%]

(d) (ii) The transition matrix is of the form considered in (b), so $\pi_1 = 0$ and the stationary distribution $\pi_2 = [\ x_1 \ \ x_2 \ \ x_3\ ]$ satisfies

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 0.0 & a & 1-a \\ 1.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

This yields

$$\begin{aligned} x_2 + x_3 &= x_1 \\ ax_1 &= x_2 \\ (1-a)x_1 &= x_3 \end{aligned}$$

leading to the following solution

$$\pi = \begin{bmatrix} 0 & \frac{1}{2} & \frac{a}{2} & \frac{1-a}{2} \end{bmatrix}$$

[20%]

(d)(iii) The equations associated with this problem are (the notation is q_i is the expected time to first visit state 4 given start in state i):

$$\begin{aligned} q_1 &= 0.5(1+q_1) + 0.5(1+q_2) \\ q_2 &= (1-a) + a(1+q_3) = 1 + aq_3 \\ q_3 &= (1+q_2) \end{aligned}$$

It is clear that from the second and third equations that

$$q_2 = \frac{1+a}{1-a}$$

The first equation gives $q_1 = 2 + q_2$, hence

$$q_1 = 2 + \frac{1+a}{1-a} = \frac{3-a}{1-a}$$

We need $q_1 = 4$. Thus this then yields $a = 1/3$. [20%]

Assessor's Comments:

87 attempts, Average mark 14.0/20, Maximum 19, Minimum 3.

A very popular question attempted by all but 4 candidates.

The question was done quite well by many but very well by few.

Most candidates did not consider carefully enough the implications of the specific form of $\mathbf{P}$ given. They overlooked the significance of the statement that the elements of $\mathbf{P}_{11}$, $\mathbf{P}_{12}$ and $\mathbf{P}_{22}$ are non-zero (as distinct from non-negative), meaning that the first m states must be transient.

In answering parts (b) and (c) many candidates stated equations without any discussion of their significance or meaning in the context of the question.

Some candidates did not normalise the components of the stationary distribution found in part (d)(ii) or normalised it in such a way that the magnitude of the 'vector' was unity rather than the sum of its components being unity.

In part (d)(iii) several candidates did not recognise that q_4 <u>must</u> be zero and set up an additional equation linking it to q_2. Whether this ended up affecting their answer depended on the solution path followed.