IMPERIAL

MATH70005

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
Summer 2025

This paper is also taken for the relevant examination for the
Associateship of the Royal College of Science

Optimisation

Date: Friday, May 2, 2025

Time: Start time 14:00 – End time 16:30 (BST)

Time Allowed: 2.5 hours

This paper has 5 Questions.

Please Answer Each Question in a Separate Answer Booklet

This is a closed book examination.

Candidates should start their solutions to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Allow margins for marking.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO DO SO

1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = 2x_1^4 + x_2^4 - 4x_1^2 x_2 + 2x_1^2 - x_2^2 + 1$$

(a) Verify that

$$(0,0), (0, \frac{1}{\sqrt{2}}), (0, -\frac{1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, 1), (-\frac{1}{\sqrt{2}}, 1)$$

are stationary points of f.

These are the only stationary points of f, but you don't need to show this. (5 marks)

(b) (i) Classify each of these stationary points as either a local minimum, local maximum, or saddle point.

(5 marks)

(ii) Determine whether each local extremum is strict or non-strict. (3 marks)

(c) (i) Show that f is coercive. (3 marks)

(ii) Determine whether f has any global extrema. If they exist, find them and prove they are indeed global. (4 marks)

(Total: 20 marks)

2. (a) Express each of the following minimisation problems in the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 ,$$

by finding a matrix $\mathbf{A}$ and a vector $\mathbf{b}$ of appropriate dimensions.

(i) $\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^\top\mathbf{Q}\mathbf{x} + 2\mathbf{f}^\top\mathbf{x} + \mathbf{f}^\top\mathbf{Q}^{-1}\mathbf{f}$, with $\mathbf{Q}\in\mathbb{R}^{n\times n}$ positive definite and $\mathbf{f}\in\mathbb{R}^n$.

(4 marks)

(ii) $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}+\mathbf{w}\|^2 + \mathbf{x}^\top\mathbf{D}\mathbf{x}$, with $\mathbf{D} = \text{diag}\,(d_1,\ldots,d_n)\in\mathbb{R}^{n\times n}$ and $d_i > 0$ for all i.

(4 marks)

(b) Let $\mathbf{A}\in\mathbb{R}^{m\times n}$, $\mathbf{b}\in\mathbb{R}^m$, and $\mathbf{L}\in\mathbb{R}^{p\times n}$ be matrices, and let $\lambda > 0$ be a positive scalar. Consider the regularised least squares problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 + \lambda\|\mathbf{L}\mathbf{x}\|^2 .$$

(i) Show that this optimization problem has a unique solution if and only if:

$$\text{Null}(\mathbf{A})\cap\text{Null}(\mathbf{L}) = \{\mathbf{0}\}$$

(6 marks)

(ii) Consider solving this optimization problem using gradient descent with a constant stepsize $\alpha > 0$:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha\nabla f(\mathbf{x}^{(k)})$$

where $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 + \lambda\|\mathbf{L}\mathbf{x}\|^2$.

Find an upper bound for the stepsize α in terms of $\mathbf{A}$ and $\mathbf{L}$ that ensures convergence of the gradient method. Express your bound as a function of a matrix eigenvalue.

(6 marks)

(Total: 20 marks)

3. (a) Consider the set

$$C := \left\{ \mathbf{x} \in \mathbb{R}^2 \colon x_1 \geq 0, x_2 \geq 0, 2x_1 + x_2 \leq 5, -3x_1 + x_2 \geq -7 \text{ and } x_2 \leq 3 \right\}$$

Show the set C is convex. (5 marks)

(b) Using the results in (a), solve the following problem:

$$\begin{cases} \max\limits_{\mathbf{x} \in \mathbb{R}^2} & x_1^4 + \frac{13}{6} x_1^2 x_2^2 + x_2^4 \\ \text{subject to} & \mathbf{x} \in C \end{cases}$$

(10 marks)

(c) Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be two nonnegative and convex functions. Assume that the following condition holds:

$$\left(f(\mathbf{y}) - f(\mathbf{x}) \right) \left(g(\mathbf{x}) - g(\mathbf{y}) \right) \leq 0 \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^2. \tag{$*$}$$

Show the function $h(\mathbf{x}) = f(\mathbf{x}) g(\mathbf{x})$ is convex. (5 marks)

(Total: 20 marks)

4. Consider the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n, t\in\mathbb{R}} \quad & \tfrac{1}{2}\|\mathbf{x}\|_2^2 + ct \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} + t\mathbf{e}, \\ & t \geq 0, \end{aligned} \tag{P}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, c \in \mathbb{R}$, and $\mathbf{e}$ is the vector of all ones. Assume in addition that the rows of $\mathbf{A}$ are linearly independent.

(a) Find a dual problem to problem (P). Do not assign a multiplier to the non-negativity constraint of t.

(8 marks)

(b) Solve the dual problem obtained in part (a).

(6 marks)

(c) Find the optimal solution of problem (P) using complementary slackness.

(6 marks)

(Total: 20 marks)

5. MASTERY QUESTION

A community living around a lake wants to maximise the yield of fish taken out of the lake. The amount of fish at a certain time is denoted by $x(t)$. The growth rate of the fish is $\kappa x(t)$, for some $\kappa > 0$, and fish is captured with a rate $u(t)x(t)$ where $u(t)$ is the control variable, which is assumed to satisfy $0 \leq u \leq u_{\max}$. The dynamics of the fish population are then given by

$$\dot{x} = (\kappa - u)x, \quad x(0) = x_o\,, \qquad \text{with } \; x_o > 0.$$

The total amount of fish obtained during a time period T is

$$J := \int_0^T u(t)x(t)dt$$

(a) Derive necessary optimality conditions (PMP) to the control u^* that maximises the yield of fish J. (6 marks)

(b) Show that the necessary conditions are satisfied by a control that only takes values on the boundary of the constraint set. How many switching times are there?

(8 marks)

(c) Determine an equation for calculating the switching time(s). (6 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2025

This paper is also taken for the relevant examination for the Associateship.

MATH60005/MATH70005

Optimisation (Solutions)

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = 2x_1^4 + x_2^4 - 4x_1^2 x_2 + 2x_1^2 - x_2^2 + 1$$

seen $\Downarrow$

(a) To find stationary points, we compute $\nabla f = 0$:

$$\frac{\partial f}{\partial x_1} = 8x_1^3 - 8x_1x_2 + 4x_1 = 4x_1(2x_1^2 - 2x_2 + 1) = 0$$

$$\frac{\partial f}{\partial x_2} = 4x_2^3 - 4x_1^2 - 2x_2 = 2(2x_2^3 - 2x_1^2 - x_2) = 0$$

Need to evaluate

$$(0,0), (0, \frac{1}{\sqrt{2}}), (0, -\frac{1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, 1), (-\frac{1}{\sqrt{2}}, 1)$$

to show that for each one of them $\nabla f = 0$.

5, A

(b) (i) The Hessian matrix is:

seen $\Downarrow$

$$H(x_1, x_2) = \begin{pmatrix} 24x_1^2 - 8x_2 + 4 & -8x_1 \\ -8x_1 & 12x_2^2 - 2 \end{pmatrix}$$

We use trace and determinant to determine whether the Hessian is positive definite at each stationary point.

- At $(0,0)$: $H = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$ Eigenvalues 4, -2 $\Rightarrow$ saddle point.
- At $(0, \frac{1}{\sqrt{2}})$: is a saddle point as the Hessian has mixed eigenvalues.
- At $(\pm\frac{1}{\sqrt{2}}, 1), (0, -\frac{1}{\sqrt{2}})$: are local minima as the Hessian is positive definite.

5, A

(ii) The minima at $(\pm\frac{1}{\sqrt{2}}, 1), (0, -\frac{1}{\sqrt{2}})$ are strict since both eigenvalues of the Hessian are positive at these points.

seen $\Downarrow$

(c) (i) For coercivity, observe that:

3, C

$$f(x_1, x_2) \geq 2x_1^4 + x_2^4 - |4x_1^2 x_2| + 2x_1^2 - x_2^2 + 1$$

unseen $\Downarrow$

and that

$$4|x_1^2 x_2| \leq 4\left(\frac{1}{2}(x_1^2)^2 + \frac{1}{2}x_2^2\right) = 2x_1^4 + 2x_2^2\,.$$

Therefore,

$$\begin{aligned} f(x_1, x_2) &\geq 2x_1^4 + x_2^4 - (2x_1^4 + 2x_2^2) + 2x_1^2 - x_2^2 + 1 \\ &= x_2^4 - 3x_2^2 + 2x_1^2 + 1 \to \infty \end{aligned}$$

as $\|(x_1, x_2)\| \to \infty$.

3, C

(ii) Since f is coercive and continuous, the only local minima are at $(\pm\frac{1}{\sqrt{2}}, 1), (0, -\frac{1}{\sqrt{2}})$. Therefore, one of these points must be a global minimum. Evaluating:

meth seen $\Downarrow$

$$f(\frac{1}{\sqrt{2}}, 1) = f(-\frac{1}{\sqrt{2}}, 1) = 0.5\,, \qquad f(0, -\frac{1}{\sqrt{2}}) = 0.75$$

Hence, $(\pm\frac{1}{\sqrt{2}}, 1)$ are global minima. There is no global maximum as f is unbounded above (by coercivity).

4, D

2. (a) (i) The minimization problem $\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{f}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{Q}^{-1}\mathbf{f}$, with $\mathbf{Q} \in \mathbb{R}^{n\times n}$ can be expressed in the form $\min |\mathbf{A}\mathbf{x} - \mathbf{b}|^2$ as follows. $\boxed{\text{seen} \Downarrow}$

Expanding $|\mathbf{A}\mathbf{x}-\mathbf{b}|^2 = (\mathbf{A}\mathbf{x}-\mathbf{b})^\top(\mathbf{A}\mathbf{x}-\mathbf{b})$ gives $\mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x} - 2\mathbf{b}^\top\mathbf{A}\mathbf{x} + \mathbf{b}^\top\mathbf{b}$.
For the two formulations to be equivalent, comparing coefficients yields $\mathbf{A}^\top\mathbf{A} = \mathbf{Q}$ and $-2\mathbf{b}^\top\mathbf{A} = 2\mathbf{f}^\top$. Since $\mathbf{Q}$ is positive definite, we can use the Cholesky decomposition to obtain $\mathbf{A}$ as the upper triangular Cholesky factor of $\mathbf{Q}$.
From $-2\mathbf{b}^\top\mathbf{A} = 2\mathbf{f}^\top$, we obtain $\mathbf{b}^\top = -\mathbf{f}^\top\mathbf{A}^{-1}$, thus $\mathbf{b} = -(\mathbf{A}^{-1})^\top\mathbf{f}$. With these definitions, we check that $\mathbf{b}^\top\mathbf{b}$ is equal to $\mathbf{f}^\top\mathbf{Q}^{-1}\mathbf{f}$:

$$\mathbf{b}^\top\mathbf{b} = (-\mathbf{f}^\top\mathbf{A}^{-1})(-(\mathbf{A}^{-1})^\top\mathbf{f}) = \mathbf{f}^\top\mathbf{Q}^{-1}\mathbf{f}\,.$$

$\boxed{4, \text{A}}$

(ii) Let us denote by $\mathbf{Q} = \mathbf{D}^{1/2} = \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$. Then, the objective function can be rewritten as $\boxed{\text{unseen} \Downarrow}$

$$\begin{aligned}\|\mathbf{x}+\mathbf{w}\|^2 + \mathbf{x}^\top\mathbf{D}\mathbf{x} &= \|\mathbf{x}+\mathbf{w}\|^2 + \|\mathbf{Q}\mathbf{x}\|^2 \\ &= \left\| \begin{bmatrix} \mathbf{x}+\mathbf{w} \\ \mathbf{Q}\mathbf{x} \end{bmatrix} \right\|^2 \\ &= \left\| \begin{bmatrix} \mathbb{I} \\ \mathbf{Q} \end{bmatrix} \mathbf{x} - \begin{bmatrix} -\mathbf{w} \\ \mathbf{0} \end{bmatrix} \right\|^2 .\end{aligned}$$

$\boxed{4, \text{B}}$

(b) (i) The objective function can be written as:

$$\|A\mathbf{x} - \mathbf{b}\|^2 + \lambda\|L\mathbf{x}\|^2 = \mathbf{x}^\top(A^\top A + \lambda L^\top L)\mathbf{x} - 2\mathbf{b}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{b}\,.$$

This is strictly convex if and only if $A^\top A + \lambda L^\top L$ is positive definite.
For any $\mathbf{x} \neq 0$:

$$\mathbf{x}^\top(A^\top A + \lambda L^\top L)\mathbf{x} = \|A\mathbf{x}\|^2 + \lambda\|L\mathbf{x}\|^2 \geq 0$$

This is strictly positive if and only if $A\mathbf{x} \neq 0$ or $L\mathbf{x} \neq 0$, which is equivalent to $\text{Null}(A) \cap \text{Null}(L) = \{0\}$.
Therefore, a unique solution exists if and only if $\text{Null}(A) \cap \text{Null}(L) = \{0\}$. $\boxed{6, \text{B}}$

(ii) The gradient of the objective is:

$$\nabla f(\mathbf{x}) = 2(A^\top A + \lambda L^\top L)\mathbf{x} - 2A^\top\mathbf{b}$$

For gradient descent with constant stepsize α:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha\nabla f(\mathbf{x}_k)$$

For convergence, we need:

$$0 < \alpha < \frac{2}{\mathcal{L}}$$

where $\mathcal{L}$ is the Lipschitz constant of the gradient. For a quadratic function this corresponds to $2\|A^\top A + \lambda L^\top L\|_2$, where the spectral norm corresponds to the largest eigenvalue of $A^\top A + \lambda L^\top L$. $\boxed{6, \text{D}}$

3. (a) Notice that the set C can be written as

$$C = H_1 \cap H_2 \cap H_3 \cap H_4 \cap H_5$$

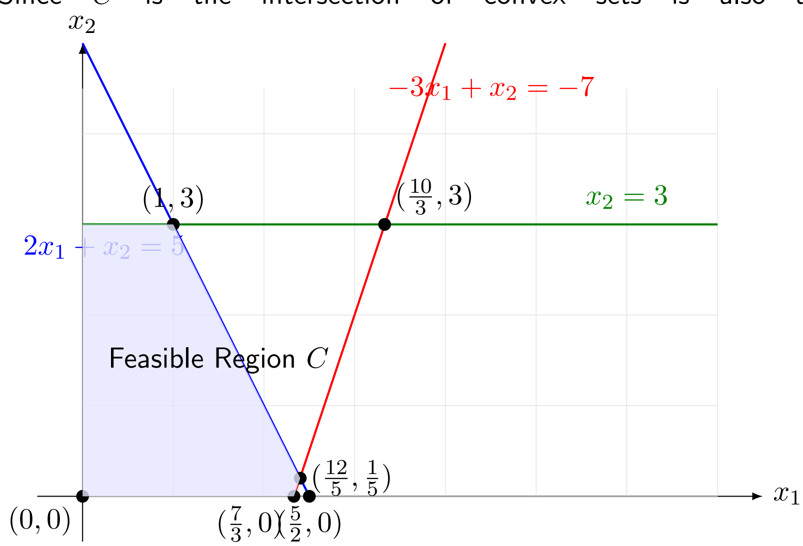
seen ⇓

where 2, A

$$\begin{aligned} H_1 &= \{\mathbf{x} \in \mathbb{R}^2 \colon x_1 \geq 0\}, \\ H_2 &= \{\mathbf{x} \in \mathbb{R}^2 \colon x_2 \geq 0\}, \\ H_3 &= \{\mathbf{x} \in \mathbb{R}^2 \colon 2x_1 + x_2 \leq 5\}, \\ H_4 &= \{\mathbf{x} \in \mathbb{R}^2 \colon -3x_1 + x_2 \geq -7\}, \\ H_5 &= \{\mathbf{x} \in \mathbb{R}^2 \colon x_2 \leq 3\}, \end{aligned}$$

and all of these sets are half-spaces. We have already proved that all half-spaces are convex. 1, A

Since C is the intersection of convex sets is also a convex set.



2, A

(b) To solve the problem we will use the theorem of *Maximum of a Convex Function Over a Compact Convex Set* from the lecture notes, which states that if the objective function is convex and the feasible set is convex and compact then at least one of the solutions of the problem is an extreme point of C. meth seen ⇓ 1, B

We have already proved that the set C is a convex polytope, and it is also easy to see that it is closed and bounded, and thus compact. 1, A

Now, we prove that the objective function meth seen ⇓

$$h(\mathbf{x}) = x_1^4 + \frac{13}{6}x_1^2x_2^2 + x_2^4,$$

is convex. To this end, we consider the Hessian of the objective function given by

$$\nabla^2 h(\mathbf{x}) = \begin{bmatrix} 12x_1^2 + \frac{13}{3}x_2^2 & \frac{26}{3}x_1x_2 \\ \frac{26}{3}x_1x_2 & \frac{13}{3}x_1^2 + 12x_2^2 \end{bmatrix}$$

and it is positive semi-definite because the trace and determinant are given by

$$\begin{aligned} \text{tr}(\nabla^2 h(\mathbf{x})) &= \frac{49}{3}x_1^2 + \frac{49}{3}x_2^2 \geq 0 \\ \det(\nabla^2 h(\mathbf{x})) &= \frac{156}{3}x_1^4 + \frac{263}{3}x_1^2x_2^2 + \frac{156}{3}x_2^2 \geq 0, \end{aligned}$$

are both non-negative. $\boxed{3, \text{B}}$

We need to find the extreme points of C which correspond to its vertices. The set C has five vertices and are the points $(0, 0)$, $(0, 3)$, $(1, 3)$, $(^{12}/_5, ^1/_5)$ and $(^7/_3, 0)$. $\boxed{\text{meth seen} \Downarrow}$

Finally, to find the solution, we need to evaluate the objective function in the different vertices and choose the one that reaches the maximum value. In this case, we obtain $h(0, 0) = 0$, $h(0, 3) = 81$, $h(1, 3) = {}^{203}/_2 \approx 101.5$, $h(^{12}/_5, ^1/_5) = {}^{6702}/_{199} \approx 33.67$ and $h(^7/_3, 0) = {}^{2401}/_{81} \approx 29.64$, and thus the solution is $\mathbf{x}^* = (1, 3)$. $\boxed{5, \text{B}}$

(c) Using the definition of convexity, we need to prove that for arbitrary points $\mathbf{x}$ and $\mathbf{y}$ belonging to $\mathbb{R}^2$ and $\lambda \in [0, 1]$ the following inequality holds $\boxed{\text{seen} \Downarrow}$

$$h(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda h(\mathbf{x}) + (1 - \lambda)h(\mathbf{y})$$

Thus, using the definition of the function h we get that

$$h(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) = f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y})g(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y})$$

Now, since the functions f, g are nonnegative and convex we obtain that

$$\begin{aligned} h(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) &\leq [\lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})]\,[\lambda g(\mathbf{x}) + (1 - \lambda)g(\mathbf{y})] \\ &= \lambda^2 f(\mathbf{x})g(\mathbf{x}) + \lambda(1 - \lambda)f(\mathbf{x})g(\mathbf{y}) \\ &\qquad + \lambda(1 - \lambda)f(\mathbf{y})g(\mathbf{x}) + (1 - \lambda)^2 f(\mathbf{y})g(\mathbf{y}) \end{aligned}$$

Now, rewriting condition $(*)$ as follows $\boxed{2, \text{A}}$

$$f(\mathbf{y})g(\mathbf{x}) + f(\mathbf{x})g(\mathbf{y}) \leq f(\mathbf{x})g(\mathbf{x}) + f(\mathbf{y})g(\mathbf{y}),$$

we obtain $\boxed{2, \text{C}}$

$$\begin{aligned} h(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) &\leq \lambda^2 f(\mathbf{x})g(\mathbf{x}) + \lambda(1 - \lambda)f(\mathbf{x})g(\mathbf{x}) \\ &\qquad + \lambda(1 - \lambda)f(\mathbf{y})g(\mathbf{y}) + (1 - \lambda)^2 f(\mathbf{y})g(\mathbf{y}) \\ &= \lambda f(\mathbf{x})g(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})g(\mathbf{y}) \\ &= \lambda h(\mathbf{x}) + (1 - \lambda)h(\mathbf{y}), \end{aligned}$$

and thus the function h is convex. $\boxed{1, \text{A}}$

4. (a) A Lagrangian of the problem excluding the non-negativity constraint is given by: meth seen ⇓

$$L(\mathbf{x}, t; \lambda) = \frac{1}{2}\|\mathbf{x}\|_2^2 + ct + \lambda^\top(\mathbf{A}\mathbf{x} - \mathbf{b} - t\mathbf{e}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \left(\mathbf{A}^\top\lambda\right)^\top \mathbf{x} + \left(c - \mathbf{e}^\top\lambda\right)t - \mathbf{b}^\top\lambda.$$

2, A

The dual objective function can be split into 2 optimisation problems

$$q(\lambda) = \min_{\mathbf{x}, t \geq 0} L(\mathbf{x}, t; \lambda) = -\mathbf{b}^\top\lambda + \min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{x}\|_2^2 + \left(\mathbf{A}^\top\lambda\right)^\top \mathbf{x}\right\} + \min_{t \geq 0} \left(c - \mathbf{e}^\top\lambda\right)t.$$

The problems consist of a quadratic optimisation and a linear problem, hence they can be solved explicity:

$$\min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{x}\|_2^2 + \left(\mathbf{A}^\top\lambda\right)^\top \mathbf{x}\right\} = -\frac{1}{2}\lambda^\top\mathbf{A}\mathbf{A}^\top\lambda, \text{ (attained at } \mathbf{x} = -\mathbf{A}^\top\lambda\Big),$$

$$\min_{t \geq 0} \left(c - \mathbf{e}^\top\lambda\right)t = \begin{cases} 0, & \mathbf{e}^\top\lambda \leq c, \\ -\infty, & \text{else,} \end{cases}$$

from where it follows that the dual objective function is

$$q(\lambda) = \begin{cases} -\frac{1}{2}\lambda^\top \mathrm{A}\mathrm{A}^\top\lambda - \mathbf{b}^\top\lambda, & \mathbf{e}^\top\lambda \leq c \\ -\infty, & \text{else.} \end{cases}$$

Thus, the dual problem is

$$\max_{\lambda} \left\{-\frac{1}{2}\lambda^\top\mathbf{A}\mathbf{A}^\top\lambda - \lambda^\top\mathbf{b} : \mathbf{e}^\top\lambda \leq c\right\}.$$

6, D

(b) We can solve it using KKT conditions, which are necessary and sufficient (convex cost and linear inequality constraint). Associating a Lagrange multiplier $\alpha \geq 0$ to the inequality constraint, the KKT conditions for the dual problem are meth seen ⇓

$$\begin{aligned} \mathrm{A}\mathrm{A}^\top\lambda + \mathbf{b} + \alpha\mathbf{e} &= 0 \\ \alpha\left(\mathbf{e}^\top\lambda - c\right) &= 0 \\ \mathbf{e}^\top\lambda &\leq c, \\ \alpha &\geq 0 \end{aligned}$$

By the first KKT equation, we have that

$$\lambda = -\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}(\mathbf{b} + \alpha\mathbf{e})$$

If $\alpha = 0$, then $\lambda = -\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{b}$. This is a solution of the KKT system, and hence an optimal dual solution, as long as $\mathbf{e}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{b} + c \geq 0$. If $\alpha > 0$, then by the complementary slackness condition we have that $\mathbf{e}^\top\lambda = c$. Substituting the expression for λ into this equation, it follows that $-\mathbf{e}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}(\mathbf{b} + \alpha\mathbf{e}) = c$, meaning that

$$\alpha = -\frac{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} + c}{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{e}}.$$

The above expression is nonnegative if and only if $\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} + c \leq 0$. We thus conclude that under the condition $\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} + c \leq 0$ the dual optimal solution is

$$\lambda = -\left(\mathbf{A}\mathbf{A}^\top\right)^{-1} (\mathbf{b} + \alpha\mathbf{e}) \text{ with } \alpha = -\frac{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} + c}{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{e}}.$$

$\boxed{5, \mathrm{A}}$

Combining the two cases, we can conclude that the optimal solution of the dual problem is

$$\lambda^* = -\left(\mathbf{A}\mathbf{A}^\top\right)^{-1} (\mathbf{b} + \alpha\mathbf{e}) \text{ with } \alpha = \frac{\left[-\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} - c\right]_+}{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{e}}.$$

$\boxed{1, \mathrm{B}}$

(c) Strong duality holds, so primal solution can recovered directly from dual solution . Denote the optimal solution of the primal problem by $(\mathbf{x}^*, t^*)$. Then

$\boxed{\text{seen} \Downarrow}$

$\boxed{2, \mathrm{A}}$

$$\mathbf{x}^* = -\mathbf{A}^\top \lambda^* = \mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} (\mathbf{b} + \alpha\mathbf{e})$$

To find t^*, substitute $(\mathbf{x}, t) = (\mathbf{x}^*, t^*)$ into the system of equality constraints in the primal problem:

$$\mathbf{A}\mathbf{x}^* = \mathbf{b} + t^*\mathbf{e}$$

Thus,

$$\mathbf{A}\mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} (\mathbf{b} + \alpha\mathbf{e}) = \mathbf{b} + t^*\mathbf{e}$$

Consequently, $\mathbf{b} + \alpha\mathbf{e} = \mathbf{b} + t^*\mathbf{e}$, meaning that $t^* = \alpha$, which also satisfies $t^* \geq 0$ as required. We conclude that the optimal solution of the primal problem is given by

$$\mathbf{x}^* = \mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} (\mathbf{b} + t^*\mathbf{e}), t^* = \frac{\left[-\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{b} - c\right]_+}{\mathbf{e}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{e}}.$$

$\boxed{4, \mathrm{C}}$

5. (a) The optimal control problem is:

$$\min_{u(\cdot)} \int_0^{\top} -ux\, dt \quad \text{subject to} \quad \dot{x} = (k-u)x, \quad x(0) = x_0$$

The Hamiltonian is:

$$\mathcal{H}(x, \lambda, u) = -ux + \lambda(k-u)x$$

The necessary conditions (PMP) are:

$$\begin{aligned} \dot{x} &= (k-u)x, \quad x(0) = x_0 \\ -\dot{\lambda} &= \frac{\partial \mathcal{H}}{\partial x} = -u + \lambda(k-u), \quad \lambda(T) = 0 \end{aligned}$$

The optimal control is:

$$u^* = \text{argmin}_{0 \le u \le u_{\max}} \mathcal{H} = \begin{cases} 0, & \lambda + 1 < 0 \\ u_{\max}, & \lambda + 1 > 0 \\ \tilde{u}, & \lambda + 1 = 0 \end{cases}$$

where $\tilde{u}$ is arbitrary in $[0, u_{\max}]$. $\boxed{6, \text{M}}$

(b) Define $\sigma = \lambda + 1$. The boundary condition on λ gives $\sigma(T) = \lambda(T) + 1 = 1 > 0$
$\Rightarrow u^*(T) = u_{\max}$
Now, $\dot{\sigma}|_{\sigma(t)=0} = \dot{\lambda}|_{\lambda+1=0} = (u - \lambda(k-u))|_{\lambda+1=0} = k > 0$
Hence, there is only one switch, since it can only pass $\sigma(t) = 0$ only once. Thus,

$$u^*(t) = \begin{cases} 0, & 0 \le t \le t' \\ u_{\max}, & t' < t \le T \end{cases}$$

for some t'. $\boxed{8, \text{M}}$

(c) Solving $\lambda(t)$ backwards for $t' \le t \le T$:

$$\lambda(t) = \frac{u_{\max}}{k - u_{\max}}(1 - e^{(k-u_{\max})(T-t)})$$

The switching time t' is determined by:

$$\lambda(t') + 1 = 0$$

$\boxed{6, \text{M}}$

Review of mark distribution:
Total A marks: 32 of 32 marks
Total B marks: 20 of 20 marks
Total C marks: 12 of 12 marks
Total D marks: 16 of 16 marks
Total marks: 100 of 80 marks
Total Mastery marks: 20 of 20 marks

MATH70005 Optimisation Markers Comments

Question 1	Most students were able to verify the stationary points by evaluating the gradient. Some mistakes when classifying stationary points, although mistakes were consistent (in the sense of a wrong eigenvalue calculation but right diagnosis). Most students identified local minima as strict. As for the coercivity calculation, very few rigorous attempts using inequalities, most of students were able to correctly identify asymptotics, but this needs to be linked to some analysis of pathological cases. Most students were able to mention coercivity when arguing for a global min, which was found by comparison among local minima.
Question 2	Most students were able to correctly cast the first LLS problem. As for the second, there is a simple answer which was reviewed in the lectures and previous coursework, however most students opted for a more complicated approach which requires to identify an additional constant in the objective, changing the cost but not the minimizer. Correct attempts for this item received full marks. For the second part of the question, the proof was discussed in class. Some students were able to correctly repeat the argument based on the definition of positive definiteness. However, a number of attempts contained circular logic errors which do not lead to the requested conclusion. For the final part, most students got the idea of a convergence bound for gradient descent, Lipschitz constant for a quadratic function, and a bound expressed in terms of the largest eigenvalue of a matrix. However, several attempts contained errors when computing the gradient of a quadratic function.
Question 3	This Question was generally answered well by majority. Q3(a, b) was a bit of a wordy question, and many candidates did not even bother sketching the diagram that was in the solution script (allocated 2 marks for part a). Though precise details were required of this sketch in part (b): candidates did this part while answering part b. So if part b had been answered well and it was evident the candidate had mastery of this, full marks were awarded. Q3(b) Most did this well and followed the approach in the solution script. A number of candidates attempted the more challenging lagrangian multipliers approach, and nearly all who followed this line of attack were unable to complete the question, due to the algebra - a reasonable proportionate mark was awarded for the attempt. Q3(c): generally done well.

Question 4 Most students didn't answer this question well. Some common mistakes include: wrong Lagrangian, include time t, assume the lagrangian multiplier is a scalar instead of a vector. Wrong dual problem, wrong optimisation of (c-e^T*lambda) when solving dual problem, no discussion of the cases a=0 and a0 (a is the lagrangian multiplier of the dual problem) not solve the time t.

Question 5 In general, students who attempted this question did the right calculations, although some mistakes were done when determining the switching structure of the problem.