IMPERIAL

MATH70005

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2024

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

Optimisation

Date: Friday, May 3, 2024

Time: 14:00 – 16:30 (BST)

Time Allowed: 2.5 hours

This paper has 5 Questions.

Please Answer Each Question in a Separate Answer Booklet

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.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

1. Consider the optimisation problem

$$\text{Minimise } f(x_1, x_2) = x_1^2 - x_1^2 x_2^2 + x_2^4$$

for $(x_1, x_2) \in \mathbb{R}^2$.

(a) Prove or disprove that f is coercive. (5 marks)

(b) Find all the stationary points of the function. (5 marks)

(c) Classify all the stationary points. Are they saddle points, strict/nonstrict local/global minima/maxima?

(10 marks)

(Total: 20 marks)

2. Consider the following set:

$$C = \left\{ \mathbf{x} \in \mathbb{R}^3 \middle| \begin{array}{l} \dfrac{x_1^2+1}{x_2} + 2x_1^2 + 5x_2^2 + 10x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \leq 7, \\ \qquad\quad x_1 \geq 0, \text{ and } x_2 \geq 1 \end{array} \right\}.$$

(a) Show that C is convex.

(8 marks)

(b) Show that the function

$$f(\mathbf{x}) = |2x_1 + 3x_2 + x_3| + x_1^2 + x_2^2 + x_3^2 + \sqrt{2x_1^2 + 4x_1x_2 + 7x_2^2 + 10x_2 + 6}$$

is convex over C.

Hint: Use the fact that a quadratic function $\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ can be rewritten as $\|\mathbf{M}(\mathbf{x} - \mathbf{x_0})\|^2 + d^2$, where x_0 is the minimiser of the quadratic function and d^2 its optimal value, and for the matrix $\mathbf{Q}$ it holds $\mathbf{Q} = \mathbf{M}\mathbf{M}^\top$. (8 marks)

(c) Consider the optimisation problem

$$\begin{cases} \text{Minimise} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in C, \end{cases}$$

where f and C stand for the function and set from the previous questions. Assume there is a local optimal solution of the problem $\mathbf{x}^*$. Is this solution a global optimiser? Is it unique? (Justify your answers). (4 marks)

(Total: 20 marks)

3. Consider the problem

$$
\text{(P)}\begin{cases} \text{Minimise} & \mathbf{x}^\top \mathbf{Q}_0 \mathbf{x} - 2\mathbf{b}^\top \mathbf{x} + c \\ \text{subject to} & \mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} \le \varepsilon, \end{cases}
$$

with $\mathbf{Q}_0, \mathbf{Q}_1 \in \mathbb{R}^{n\times n}$ positive definite matrices, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$ and $\varepsilon > 0$.

(a) Show this problem has at least one global optimal solution. (2 marks)

(b) State the KKT conditions of this problem. (2 marks)

(c) Are the KKT conditions necessary and/or sufficient? (Justify your answer). (5 marks)

(d) Let $\mathbf{y} \in \mathbb{R}^n$, find the orthogonal projection $\mathbb{P}_C(\mathbf{y})$ where

$$
C = \left\{ \mathbf{x} \in \mathbb{R}^n \colon \mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} \le \varepsilon \right\} .
$$

Hint: The solution should depend on a single parameter that is the root of a strictly decreasing continuous one-dimensional function.

(8 marks)

(e) Using your previous results, write the generic iteration of the projected gradient method for problem (P). Write explicitly a step-size rule, and give a suitable stopping criterion.

(3 marks)

(Total: 20 marks)

4. Consider the problem

$$\begin{cases} \text{Minimise} & \alpha \\ \text{subject to} & \begin{aligned} \mathbf{A}\mathbf{x} &= \alpha\mathbf{f}, \\ x &\in X \end{aligned} \end{cases}$$

where $\alpha \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{f} \in \mathbb{R}^m$, $\varepsilon > 0$ and the set $X = \{\mathbf{x} \in \mathbb{R}^n \colon \mathbf{x}^\top \mathbf{x} \le \varepsilon\}$. Assume that the rows of $\mathbf{A}$ are linearly independent.

(a) Does the strong duality hold for this problem? (Justify your answer). (4 marks)

(b) Compute the dual objective and formulate the dual problem. (6 marks)

(c) Solve the dual problem. (6 marks)

(d) Use the completementary slackness conditions to recover the solution of the primal problem. (4 marks)

(Total: 20 marks)

5. MASTERY QUESTION

(a) We want to find a control signal $u(t)$ on the interval $0 \le t \le 1$ that takes the system

$$\dot{x} = u$$

from $x(0) = 1$ to $x(1) = 0$ while minimising

$$\int_0^1 \frac{u(t)^2}{2} dt\,.$$

Show that the necessary conditions for optimality are satisfied by a constant u value and compute it.

(10 marks)

(b) The system

$$\dot{x} = -x + u$$

subject to $|u(t)| \le 1$ for all $0 \le t \le 1$, is to be controlled so that $x(1) = 0$, and the cost

$$\int_0^1 |u(t)| dt$$

is minimised. A possible control is

$$u(t) = \begin{cases} 0 & 0 \le t < 0.5, \\ -1 & 0.5 \le t \le 1. \end{cases}$$

Show that this control satisfies the necessary conditions for optimality given some value of $x(0)$ (which is not necessary to find).

(10 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2024

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

MATH60005/70005

Optimisation (Solutions)

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

1. (a) The function $f(x_1, x_2) = x_1^2 - x_1^2 x_2^2 + x_2^4$ is not coercive. It is enough to notice that $\boxed{\text{seen} \Downarrow}$

$$f(x_1, 2) = 16 - 3x_1^2 \to -\infty \quad \text{for } x_1 \to +\infty$$

$\boxed{5, \text{B}}$

(b) To find the stationary points, we compute the gradient $\boxed{\text{seen} \Downarrow}$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 2x_1 x_2^2 \\ -2x_1^2 x_2 + 4x_2^3 \end{bmatrix}$$

Solving the system $\nabla f(x_1, x_2) = [0, 0]^\top$, we find five stationary points.

$$\mathbf{x}^* = [0, 0]^\top$$
$$\mathbf{x}^\dagger = [\pm\sqrt{2}, \pm 1]^\top$$

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

(c) To classify the stationary points we consider the Hessian of the function

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 - 2x_2^2 & -4x_1 x_2 \\ -4x_1 x_2 & 12x_2^2 - 2x_1^2 \end{bmatrix}$$

Now, we need to study if the Hessian at the stationary points is positive, negative definite or indefinite matrix. For the stationary points $\mathbf{x}^\dagger = [\pm\sqrt{2}, \pm 1]^\top$, we can write the Hessian in terms of $\alpha, \beta \in \{-1, 1\}$, as follows

$$\nabla^2 f(\alpha\sqrt{2}, \beta) = \begin{bmatrix} 0 & -4\sqrt{2}\alpha\beta \\ -4\sqrt{2}\alpha\beta & 8 \end{bmatrix}$$

Thus, tr $\left(\nabla^2 f(\alpha\sqrt{2}, \beta)\right) = 8$ and det $\left(\nabla^2 f(\alpha\sqrt{2}, \beta)\right) = -32$, and therefore this matrix is indefinite and the four points are saddle points. $\boxed{4, \text{A}}$

On the other hand,

$$\nabla^2 f(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

which is a positive semidefinite matrix and thus the point is either a saddle point or a local minimum. Notice that

$$f(0, 0) = 0 \leq f(x_1, x_2) = x_1^2(1 - x_2^2) + x_2^4$$

for $|x_2| \leq 1$, and thus it is a local minimum. $\boxed{3, \text{B}}$

However, it is not a global minimum since the objective function is not coercive. $\boxed{3, \text{A}}$

2. (a) Notice that the set C can be rewritten as $C = \widetilde{C} \cap H_1^+ \cap H_2^+$ where seen $\Downarrow$

$$H_1^+ = \{(x_1, x_2, x_3) \colon (x_1, x_2, x_3)^\top (1, 0, 0) \geq 0\},$$
$$H_2^+ = \{(x_1, x_2, x_3) \colon (x_1, x_2, x_3)^\top (0, 1, 0) \geq 1\},$$

which are both convex as they are half-spaces. Now, we focus on the set 2, C

$$\widetilde{C} = \{(x_1, x_2, x_3) \colon g(x_1, x_2, x_2) \leq 7\}$$

which is the level set of the function $g(x_1, x_2, x_2) = \frac{x_1^2+1}{x_2} + 2x_1^2 + 5x_2^2 + 10x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3$. We will prove g is convex over $H_1^+ \cap H_2^+$. This function in turn can be rewritten as the sum of two functions 2, C

$$g_1(x_1, x_2) = \frac{x_1^2 + 1}{x_2}, \text{with } \nabla^2 g_1(x_1, x_2) = \begin{bmatrix} 2/x_2 & -2x_1/x_2^2 \\ -2x_1/x_2^2 & 2(x_1^2+1)/x_2^3 \end{bmatrix}$$

The trace of the Hessian is $\operatorname{tr}\left(\nabla^2 g_1(x_1, x_2)\right) = 2/x_2 + 2(x_1^2 + 1)/x_2^3 > 0$ over $H_1^+ \cap H_2^+$ and

$$\det \begin{bmatrix} 2/x_2 & -2x_1/x_2^2 \\ -2x_1/x_2^2 & 2(x_1^2+1)/x_2^3 \end{bmatrix} = \frac{4}{x_2^4} > 0.$$

Therefore g_1 is strictly convex as its Hessian is positive definite. 2, B

Finally, the function g_2 is convex, since it can be rewritten as

$$g_2(x_1, x_2, x_3) = [x_1, x_2, x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

and the associated matrix is positive definite, as all its principal minors are positive.

$$q_{11} = 2 > 0 \quad \det \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} = 6 > 0, \quad \det \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 10 \end{bmatrix} = 57 > 0$$

2, B

(b) Notice that the function seen $\Downarrow$

$$f(x_1, x_2, x_3) = |2x_1 + 3x_2 + x_3| + x_1^2 + x_2^2 + x_3^2 + \sqrt{2x_1^2 + 4x_1x_2 + 7x_2^2 + 10x_2 + 6}$$

can be rewritten as the sum of three functions

$$f_1(x_1, x_2, x_3) = |2x_1 + 3x_2 + x_3|$$
$$f_2(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$$
$$f_3(x_1, x_2, x_3) = \sqrt{2x_1^2 + 4x_1x_2 + 7x_2^2 + 10x_2 + 6}$$

We will prove each one is convex and therefore f will be convex as the sum of convex functions $\boxed{1, \text{A}}$

In the case of f_1, notice that is the composition of an affine functional $[2,3,1]^\top[x_1,x_2,x_3]$ and the absolute value which is convex. $\boxed{2, \text{A}}$

Now, let us consider $f_2 = \|[x_1,x_2,x_3]\|^2$ which convex as it is a norm. $\boxed{1, \text{A}}$

Finally, let us focus on $\boxed{\text{unseen} \Downarrow}$

$$f_3(x_1,x_2,x_3) = \sqrt{2x_1^2 + 4x_1x_2 + 7x_2^2 + 10x_2 + 6}$$

The function inside of the squared root can be written as

$$\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$$

with $\mathbf{Q} = \begin{bmatrix} 2 & 2 \\ 2 & 7 \end{bmatrix}$, $\mathbf{b} = [0, 10]$ and $c = 6$. The matrix $\mathbf{Q}$ is positive definite and therefore there exists $\mathbf{M}$ such that $\mathbf{Q} = \mathbf{M}\mathbf{M}^\top$. Using the hint,

$$f_3(x_1,x_2,x_3)) = \sqrt{\|\mathbf{M}(\mathbf{x}-\mathbf{x_0})\|^2 + d^2} = \left\| \begin{bmatrix} \mathbf{M}(\mathbf{x}-\mathbf{x_0}) \\ d \end{bmatrix} \right\|$$

which is an affine transformation composed with the norm which is convex. Therefore, we obtain the convexity of f_3 and thus of f. $\boxed{4, \text{B}}$

(c) Since we assume there is a local optimal solution of a convex optimisation problem, we know it is global. To guarantee the uniqueness of the solution, we should prove the objective function is strictly convex. Since all three functions are strictly convex, then indeed the local optimizer is the only global solution of the problem. $\boxed{\text{seen} \Downarrow}$ $\boxed{4, \text{A}}$

3. (a) We start by noticing that the objective function is continuous since it is quadratic. Moreover, the constraint represents an ellipsoid and we know it is compact. Therefore, using the Weiestrass theorem we know there exists at least one optimal global solution. seen ⇓

2, A

(b) We start by considering the Lagrangian of the problem meth seen ⇓

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^\top \mathbf{Q}_0 \mathbf{x} - 2\mathbf{b}^\top \mathbf{x} + c + \lambda \left(\mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} - \varepsilon \right)$$

Then, the KKT conditions are given by the following system

$$\begin{aligned} \mathbf{Q}_0 \mathbf{x} - \mathbf{b} + \lambda \mathbf{Q}_1 \mathbf{x} &= 0 \\ \lambda \left(\mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} - \varepsilon \right) &= 0 \\ \lambda &\geq 0 \end{aligned}$$

2, A

(c) Since the constraint is nonlinear and convex, we need to check if Slater's condition is satisfied. 1, A

In this case, since $\varepsilon > 0$ it is enough to take $\hat{\mathbf{x}} = \mathbf{0} \in \mathbb{R}^n$ since $\hat{\mathbf{x}}^\top \mathbf{Q}_1 \hat{\mathbf{x}} = 0 < \varepsilon$, and thus, the KKT conditions are necessary. 2, B Since the objective function is convex, the KKT conditions are sufficient.

2, A

(d) By definition, to find the orthogonal projection over C, we have to solve the following problem

$$\begin{cases} \text{Minimise} & \|\mathbf{x} - \mathbf{y}\|^2 \\ \text{subject to} & \mathbf{x} \in C. \end{cases}$$

seen ⇓

To solve this problem, we use the KKT conditions given by the following system 2, C

$$\begin{aligned} \mathbf{x} - \mathbf{y} + \lambda \mathbf{Q}_1 \mathbf{x} &= 0 \\ \lambda(\mathbf{x}^\top \mathbf{Q}_1 \mathbf{x} - \varepsilon) &= 0 \end{aligned}$$

where $\lambda \geq 0$. 1, C

We analyse by cases. If $\lambda = 0$, then $\mathbf{x}^* = \mathbf{y}$, which happens only if $\mathbf{y} \in C$. Otherwise, if $\lambda > 0$, then we have that

$$\mathbf{x}(\lambda) = (\mathbb{I} + \lambda \mathbf{Q}_1)^{-1} \mathbf{y}$$

where $\mathbb{I}$ is the identity matrix of dimension $n \times n$, and the inverse exists since $\mathbf{Q}_1$ is positive definite and $\lambda > 0$. 1, C

Using the complementary slackness conditions, since $\lambda > 0$, it must hold

$$(\mathbf{x}(\lambda))^\top \mathbf{Q}_1 \mathbf{x}(\lambda) = \varepsilon.$$

unseen ⇓

Thus, we need to find λ such that is the root of the one-dimensional function

$$g(\lambda) = (\mathbf{x}(\lambda))^\top \mathbf{Q}_1 \mathbf{x}(\lambda) - \varepsilon$$

Notice that this root exists since, $g(0) = \mathbf{y}^\top \mathbf{Q}_1 \mathbf{y} - \varepsilon > 0$, since $\mathbf{y} \notin C$, and $g(\lambda) \to -\varepsilon < 0$ when $\lambda \to +\infty$, and it is strictly decreasing. 2, C If we call λ^* the root of g, then the orthogonal projection is given by

$$\mathbf{x}^* = (\mathbb{I} + \lambda^* \mathbf{Q}_1)^{-1} \mathbf{y}.$$

2, C

(e) The gradient of the objective function is given by $\boxed{\text{seen} \Downarrow}$

$$2\mathbf{Q}_0\mathbf{x} - 2\mathbf{b}$$

One can use a constant step-size rule using the Lipschitz constant of the gradient, which in this case is given by $L = \|\mathbf{Q}_1\|$, explicitly this means, we use

$$t = \frac{1}{L} = \frac{1}{\|\mathbf{Q}_1\|}$$

Then, a generic iteration of the projected gradient algorithm is given by $\boxed{1, \text{A}}$

$$\mathbf{x}^{k+1} = \mathbf{P}_C(\mathbf{x}^k - t(2\mathbf{Q}_0\mathbf{x}^k - 2\mathbf{b}))$$

A suitable stopping criterion is $\boxed{1, \text{A}}$

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| < \text{tol}.$$

$\boxed{1, \text{A}}$

4. (a) The objective function α is convex and we have linear equality constraints with a convex set X, then strong duality holds. seen $\Downarrow$

4, A

(b) To compute the dual objective, we formulate the Lagrangian of the problem. meth seen $\Downarrow$

$$\mathcal{L}(\alpha, \mathbf{x}, \mu) = \alpha + \mu^\top(\mathbf{A}\mathbf{x} - \alpha\mathbf{f})$$

Then, the dual objective is given by

$$q(\mu) = \min_{(\alpha,\mathbf{x})\in\mathbb{R}\times X} \alpha(1 - \mu^\top) + \mu^\top \mathbf{A}\mathbf{x}$$

since the objective function is separable, we can solve two indepedent problems for $\alpha \in \mathbb{R}$ and for $\mathbf{x} \in X$.

The first problem is then

$$\min_{\alpha\in\mathbb{R}} \alpha(1 - \mu^\top \mathbf{f}) = \begin{cases} 0, & \text{if } \mu^\top \mathbf{f} = 1, \\ -\infty, & \text{otherwise.} \end{cases}$$

On the other hand, the second problem is given by 2, D

$$\min_{\mathbf{x}\in X} (\mathbf{A}^\top \mu)\mathbf{x}$$

using the Cauchy-Schwarz inequality one can bound from below the objective function and obtain

$$(\mathbf{A}^\top \mu)\mathbf{x} \geq -\|\mathbf{A}^\top \mu\|\|\mathbf{x}\| \geq -\|\mathbf{A}^\top \mu\|\sqrt{\varepsilon}.$$

This bound is attained when

$$\mathbf{x}^* = -\sqrt{\varepsilon}\frac{\mathbf{A}^\top \mu}{\|\mathbf{A}^\top \mu\|}.$$

The dual objective is then 2, D

$$q(\mu) = \begin{cases} -\sqrt{\varepsilon}\|\mathbf{A}^\top \mu\|, & \text{if } \mu^\top \mathbf{f} = 1, \\ -\infty, & \text{otherwise.} \end{cases}$$

seen $\Downarrow$

The dual problem is given by

$$\begin{cases} \text{Maximise} & -\sqrt{\varepsilon}\|\mathbf{A}^\top \mu\| \\ \text{subject to} & \mu^\top \mathbf{f} = 1. \end{cases}$$

2, D

(c) To solve the dual problem, we start by noticing that the solution of the dual problem will be the same as the solution of the following problem seen $\Downarrow$

$$\begin{cases} \text{Maximise} & -\varepsilon\mu^\top \mathbf{A}\mathbf{A}^\top \mu \\ \text{subject to} & \mu^\top \mathbf{f} = 1. \end{cases}$$

Thus, we use the KKT-conditions to solve this problem. The Lagrangian is given by 1, D

$$\mathcal{L}_q(\mu, \eta) = -\varepsilon\mu^\top \mathbf{A}\mathbf{A}^\top \mu + \eta(\mu^\top \mathbf{f} - 1)$$

Then, we need to solve the system

$$\begin{aligned} -2\varepsilon \mathbf{A}\mathbf{A}^\top \mu + \eta \mathbf{f} &= 0 \\ \mu^\top \mathbf{f} - 1 &= 0. \end{aligned}$$

From the first equation we get that

$$\mu^* = \frac{\eta}{2\varepsilon}(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}$$

which can be done since the rows of $\mathbf{A}$ are linearly independent and therefore $\mathbf{A}\mathbf{A}^\top$ is positive definite (and thus invertible). Now, using the fact that $\mu^\top \mathbf{f} - 1$, we obtain that

$$\eta^* = \frac{2\varepsilon}{\mathbf{f}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}}$$

and

1, D

$$\mu^* = \frac{1}{\mathbf{f}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}}(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}$$

and the optimal value is given by

2, D

$$q^* = -\sqrt{\frac{\varepsilon}{\mathbf{f}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}}}$$

2, D

(d) The complementary slackness conditions on duality state that

seen ⇓

$$\mathbf{x}^* = -\sqrt{\varepsilon}\frac{\mathbf{A}^\top \mu^*}{\|\mathbf{A}^\top \mu^*\|} = -\frac{\sqrt{\varepsilon}}{\|\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)\mathbf{f}\|}\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}$$

Finally, since strong complementary holds, we have that

2, D

$$\alpha^* = q^* = -\sqrt{\frac{\varepsilon}{\mathbf{f}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{f}}}$$

2, D

5. (a) We write the Hamiltonian: $H(u, \lambda) = \frac{u^2}{2} + \lambda u$. The PMP reads $\boxed{2, \text{M}}$

$$\dot{\lambda} = 0\,, \qquad \lambda(1) = \mu,$$

for some value of μ . Since, give λ, the optimal control minimizes the Hamiltonian, it follows that $u^* = -\lambda(t) = -\mu$. Therefore, the optimal trajectory is a straight line $\boxed{2, \text{M}}$ $\boxed{2, \text{M}}$

$$x^*(t) = -\mu t + C_1, \qquad x(0) = 1, x(1) = 0,$$

leading to $C_1 = \mu = 1$ and to $u^* = -1$. $\boxed{2, \text{M}}$

(b) Setting: Hamiltonian $H(x, u, \lambda) = |u| + \lambda(u - x)$, $\phi(x) = 0$ and $\psi(x) = x$. PMP: $\dot{\lambda} = \lambda$, $\lambda(1) = \phi_x(x(1)) + \mu\psi_x(x(1)) = \mu$. From PMP we obtain $\lambda(t) = \mu e^{t-1}$, which is monotone. . Since the Hamiltonian $H = (sign(u) + \lambda)u = \sigma u$ is minimized at all times, it follows that: $\boxed{2, \text{M}}$ $\boxed{2, \text{M}}$ $\boxed{2, \text{M}}$ $\boxed{2, \text{M}}$

$$\begin{cases} \lambda < -1 & u = 1, \\ -1 \le \lambda \le 1 & u = 0, \\ \lambda > 1 & u = -1 \end{cases}$$

Given the control candidate, going from 0 to -1, if we select μ so that λ goes from < 1 to ≥ 1 at $t = 0.5$, that is $\mu = e^{1/2}$, then the proposed control satisfies the necessary conditions. $\boxed{2, \text{M}}$ $\boxed{2, \text{M}}$

Review of mark distribution:

Total A marks: 34 of 32 marks

Total B marks: 18 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

MATH60005 Optimisation

Question	Marker's comment
1	(a) Most students were able to identify trajectories to show that the function is not coercive. A few students showed coercivity along a single trajectory, which is not sufficient.(b) Most students had no problems in finding 5 stationary points by solving grad f=0.(c) Classification for 4 saddle points was done correctly by most students. For the point (0,0), most students were able to identify either a local min or a saddle point. Fewer students correctly showed a local min by using the definition and discarding a global min because of the lack of coercivity.
2	(a)Most students considered the set C as the level set of a function which will be convex if the function in question is convex. However, only a few students realised that C is indeed the intersection of the level set of a function and the hyperplanes x_1 gt;= 0 and x_2 gt;= 1 which are also convex. These students used the fact that the intersection of convex sets is also a convex set. To prove the level set of a function is convex, most students proved the corresponding function is convex. Few students proved the function is convex without mentioning the level set of the function, which is wrong. To prove the associated function is convex, most students considered the different summands and used the fact that the sum of convex functions is convex. For the first one, some students used the result studied in the lectures that a quad-over-lin function is convex. Some students also computed the Hessian of this function and proved it is positive definite. For the second summand, most students used Sylvester’s criterion to check if the matrix is positive definite and thus concluded the function is convex. Few students computed only the determinant and the trace of the matrix which is not enough to conclude such a matrix (3X3) is positive definite.(b)Most students proved the convexity of the function by considering the different summands and proving each term is convex. Most students correctly concluded that the first summand is convex since it is the composition of an affine linear function with a convex function (the absolute value). Few students confused this term with a “norm” and directly concluded this is convex, which is not enough. Most students used the fact that the second term is the norm in R^3 and thus concluded it is convex. Some students computed the Hessian of this term and concluded the function is convex since its Hessian is positive definite. Most students were not able to prove the last term is convex. Only a few students used the hint correctly and rewrote the quadratic function as suggested. Few students proved the function sqrt(t^2 + c^2) is convex and thus the third term is the composition of an affine linear function with the norm which is convex and then with the convex and strictly increasing function. Some students tried to use the fact that sqrt(t) is convex which is false. Few students used the fact that the function can be understood as the norm of an extended vector.(c)Some students used the fact that for convex optimisation problems local optimality implies global optimality. Some students considered the KKT conditions and in some cases even strong duality conditions for this problem; however, it is not enough to guarantee global optimality. Few students mentioned that the uniqueness is obtained from the strict convexity of the objective function and correctly concluded that it is indeed strictly convex, and thus obtained the uniqueness of the solution.

MATH60005 Optimisation

Question	Marker's comment
3	(a)Most students tried to use the fact that a convex objective function over a convex set is enough to guarantee the existence of a minimiser which is not. Few students used the fact that the objective function is continuous, and the constraints form a compact set and the Weierstrass Theorem to guarantee the existence of solutions. Few students use the fact that the function is coercive to guarantee existence.(b)Most students correctly stated the KKT conditions. Few students made mistakes in the computation or did not explicitly state the gradient of the Lagrangian.(c)Most students used the fact that for convex optimisation problems the KKT conditions are sufficient. Some students considered that for optimisation problems with nonlinear constraints, KKT conditions are necessary only if Slater’s condition is satisfied and showed a vector which satisfies this condition. Few students confuse the necessary with sufficient conditions.(d)Most students formulated correctly the orthogonal projection problem and stated its corresponding KKT conditions. Most students considered the case when the Lagrange multiplier is zero and concluded that this implies the orthogonal projection is exactly the point of evaluation and this holds only if this point belongs to C. Only a few students considered the case when the Lagrange multiplier is positive and defined the one-dimensional function depending on this multiplier. Very few analysed why this function has a root and expressed the solution in terms of the root of this one-dimensional function.(e)Most students wrote the generic iteration of the project gradient method correctly. Regarding the step size rule, different solutions were presented. Some students considered a constant step size without further information or equal to one, which does not guarantee the convergence of the algorithm. Some students correctly stated that the step size can be considered constant as long as it equals the Lipschitz constant of the objective function’s gradient. Some students also considered the exact line search and stated its associated condition. Few students considered a projected backtracking method to compute the step size. Most students considered the norm of the difference between two consecutive iterations as the stopping criterion.
4	(a)The majority of the students correctly identified the cost and claim that convex cost with convex constraints leads to strong duality. Some students did not properly justify why the cost was convex. Constraints were justified in more detail, some students also used Slater’s condition.(b)Students that worked over the correct cost obtained different answers that qualitatively approximate the right answer. Some students opted to include a multiplier for the ball constraint, and in many cases this was correctly executed leading to the same answer asconsidering it as part of the minimization and using Cauchy-Schwartz (some students still were able to get this same procedure). Most students were able to identify two separate optimization problems as part of the duality calculation.(c)For a correct dual formulation, the answer can be obtained by direct computation. However, many students opted for solving using KKT conditions, which in some cases were correctly cast and obtained the same solution. Most of the errors at this stage can be attributed to a wrong dual problem in (b).(d)Students who solved (b) and (c) were able to retrieve the primal variables directly, most students making errors in previous steps only stated general complementary slackness conditions.

MATH70005 Optimisation

Question	Marker's comment
1	(a) Most students were able to identify trajectories to show that the function is not coercive. A few students showed coercivity along a single trajectory, which is not sufficient.(b) Most students had no problems in finding 5 stationary points by solving grad f=0.(c) Classification for 4 saddle points was done correctly by most students. For the point (0,0), most students were able to identify either a local min or a saddle point. Fewer students correctly showed a local min by using the definition and discarding a global min because of the lack of coercivity.
2	(a)Most students considered the set C as the level set of a function which will be convex if the function in question is convex. However, only a few students realised that C is indeed the intersection of the level set of a function and the hyperplanes x_1 gt;= 0 and x_2 gt;= 1 which are also convex. These students used the fact that the intersection of convex sets is also a convex set. To prove the level set of a function is convex, most students proved the corresponding function is convex. Few students proved the function is convex without mentioning the level set of the function, which is wrong. To prove the associated function is convex, most students considered the different summands and used the fact that the sum of convex functions is convex. For the first one, some students used the result studied in the lectures that a quad-over-lin function is convex. Some students also computed the Hessian of this function and proved it is positive definite. For the second summand, most students used Sylvester’s criterion to check if the matrix is positive definite and thus concluded the function is convex. Few students computed only the determinant and the trace of the matrix which is not enough to conclude such a matrix (3X3) is positive definite.(b)Most students proved the convexity of the function by considering the different summands and proving each term is convex. Most students correctly concluded that the first summand is convex since it is the composition of an affine linear function with a convex function (the absolute value). Few students confused this term with a “norm” and directly concluded this is convex, which is not enough. Most students used the fact that the second term is the norm in R^3 and thus concluded it is convex. Some students computed the Hessian of this term and concluded the function is convex since its Hessian is positive definite. Most students were not able to prove the last term is convex. Only a few students used the hint correctly and rewrote the quadratic function as suggested. Few students proved the function sqrt(t^2 + c^2) is convex and thus the third term is the composition of an affine linear function with the norm which is convex and then with the convex and strictly increasing function. Some students tried to use the fact that sqrt(t) is convex which is false. Few students used the fact that the function can be understood as the norm of an extended vector.(c)Some students used the fact that for convex optimisation problems local optimality implies global optimality. Some students considered the KKT conditions and in some cases even strong duality conditions for this problem; however, it is not enough to guarantee global optimality. Few students mentioned that the uniqueness is obtained from the strict convexity of the objective function and correctly concluded that it is indeed strictly convex, and thus obtained the uniqueness of the solution.

MATH70005 Optimisation

Question Marker's comment

3 (a)Most students tried to use the fact that a convex objective function over a convex set is enough to guarantee the existence of a minimiser which is not. Few students used the fact that the objective function is continuous, and the constraints form a compact set and the Weierstrass Theorem to guarantee the existence of solutions. Few students use the fact that the function is coercive to guarantee existence.(b)Most students correctly stated the KKT conditions. Few students made mistakes in the computation or did not explicitly state the gradient of the Lagrangian.(c)Most students used the fact that for convex optimisation problems the KKT conditions are sufficient. Some students considered that for optimisation problems with nonlinear constraints, KKT conditions are necessary only if Slater’s condition is satisfied and showed a vector which satisfies this condition. Few students confuse the necessary with sufficient conditions.(d)Most students formulated correctly the orthogonal projection problem and stated its corresponding KKT conditions. Most students considered the case when the Lagrange multiplier is zero and concluded that this implies the orthogonal projection is exactly the point of evaluation and this holds only if this point belongs to C. Only a few students considered the case when the Lagrange multiplier is positive and defined the one-dimensional function depending on this multiplier. Very few analysed why this function has a root and expressed the solution in terms of the root of this one-dimensional function.(e)Most students wrote the generic iteration of the project gradient method correctly. Regarding the step size rule, different solutions were presented. Some students considered a constant step size without further information or equal to one, which does not guarantee the convergence of the algorithm. Some students correctly stated that the step size can be considered constant as long as it equals the Lipschitz constant of the objective function’s gradient. Some students also considered the exact line search and stated its associated condition. Few students considered a projected backtracking method to compute the step size. Most students considered the norm of the difference between two consecutive iterations as the stopping criterion.

4 (a)The majority of the students correctly identified the cost and claim that convex cost with convex constraints leads to strong duality. Some students did not properly justify why the cost was convex. Constraints were justified in more detail, some students also used Slater’s condition.(b)Students that worked over the correct cost obtained different answers that qualitatively approximate the right answer. Some students opted to include a multiplier for the ball constraint, and in many cases this was correctly executed leading to the same answer as considering it as part of the minimization and using Cauchy-Schwartz (some students still were able to get this same procedure). Most students were able to identify two separate optimization problems as part of the duality calculation.(c)For a correct dual formulation, the answer can be obtained by direct computation. However, many students opted for solving using KKT conditions, which in some cases were correctly cast and obtained the same solution. Most of the errors at this stage can be attributed to a wrong dual problem in (b).(d)Students who solved (b) and (c) were able to retrieve the primal variables directly, most students making errors in previous steps only stated general complementary slackness conditions.

5 (a)And (b) Most students were able to correctly derive Pontryagin’s conditions and from here to solve the exercises. No significant conceptual errors.