MATH70005

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

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

Optimisation

Date: 9 May 2023

Time: 14:00 – 16:30 (BST)

Time Allowed: 2.5hrs

This paper has 5 Questions.

Please Answer Questions 1-2, and Questions 3-5 in Separate Answer Booklets

Candidates should start their answers 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. (a) Consider the quadratic function

$$f(x, y) = x^2 + \alpha xy + y^2 + x + \beta y\,,$$

with parameters $\alpha, \beta \in \mathbb{R}$.

(i) For each value of α and β, find the set of all stationary points. (7 marks)

(ii) Which of these stationary points are global minima? (5 marks)

(b) Consider the function

$$f(x, y) = x^2 - 2xy^2 + \frac{1}{2}y^4\,.$$

(i) Is this function coercive? Justify your answer. (3 marks)

(ii) Find all the stationary points and classify them. (5 marks)

(Total: 20 marks)

2. (a) Consider the composition $h := f \circ g : \mathbb{R}^2 \to \mathbb{R}$ where

$$g(x, y) = x^2 - xy + y^2 - 2x + y\,,$$

and $f(t) = e^t$.

(i) Is h convex? Justify your answer. (5 marks)

(ii) Find a minimizer for

$$\begin{aligned} \min_{x,y} \quad & -h(x, y) \\ \text{s.t.} \quad & x - y \leq 0 \\ & y + x \leq 1 \\ & 0 \leq x\,. \end{aligned}$$

(8 marks)

(b) Suppose we apply gradient descent to minimise

$$f(x) = \frac{2}{3}|x|^3 + \frac{1}{2}x^2\,,$$

with stepsize given by $t^k = \frac{1}{k+1}$. Show that for $|x^0| \geq 1$ the method diverges.
Hint: analyse what happens with the iteration if $|x^k| \geq k + 1$. (7 marks)

(Total: 20 marks)

3. Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$ and consider the set

$$C_{\mathbf{l},\mathbf{u}} = \{\mathbf{x} \in \mathbb{R}^n : l_i \leq x_i \leq u_i\,, \quad i = 1, \ldots, n\}\,,$$

a function $f : \mathbb{R}^n \to \mathbb{R}$ that is continuously differentiable over C, and the minimisation problem

$$\min_{\mathbf{x}} f(\mathbf{x})\,, \quad \text{subject to } \mathbf{x} \in C\,. \tag{P}$$

(a) Without making any further assumption about f, and using the definition of stationarity in convex optimisation, show that if a point $\mathbf{x}^*$ satisfies

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0, & l_i < x_i^* < u_i \\ \leq 0, & x_i^* = u_i \\ \geq 0, & x_i^* = l_i \end{cases}$$

is a stationary point of problem (P). (6 marks)

(b) Let $f(\mathbf{x}) := \|\mathbf{x} - \mathbf{b}\|_2^2$, with $\mathbf{b} \in \mathbb{R}^n$. Show that a minimiser of problem (P) verifies the stationarity characterization from part (a) (you need to show this explicitly). (8 marks)

(c) Suppose that you have a routine that, given a vector $\mathbf{b} \in \mathbb{R}^n$, computes the solution to (P) for $f(\mathbf{x}) := \|\mathbf{x} - \mathbf{b}\|_2^2$ as above. Explain how can you use this routine to construct an iterative method to solve

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{c}\|_2^2\,, \quad \text{subject to } \mathbf{x} \in C\,, \tag{P}$$

where $A \in \mathbb{R}^{n \times n}, A \succ 0$, and $\mathbf{c} \in \mathbb{R}^n$. Write a generic iteration for this method. (6 marks)

(Total: 20 marks)

4. (a) Consider the convex optimisation problem

$$\begin{aligned} &\min \; f(x,y) := x \\ \text{s.t.} \quad & x^2 + (y-1)^2 \leq 1\,, \\ & x^2 + (y+1)^2 \leq 1\,. \end{aligned}$$

(i) Show that there exists a unique global minimum. (3 marks)

(ii) Does the global minimum satisfy the KKT conditions? Why does this happen? (4 marks)

(b) Consider the set

$$C = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + 2x_2^2 + 3x_3^2 \leq 1\}\,,$$

(i) Is C convex a convex set? Justify your answer. Write the orthogonal projection onto C as an optimisation problem. Does this problem have unique solution? (4 marks)

(ii) Using convex optimisation tools, find a formula for the orthogonal projection of a vector $\mathbf{y} \in \mathbb{R}^3$ onto C. (9 marks)

(Total: 20 marks)

5. MASTERY QUESTION

(a) A population of fish in a lake evolves according to the equation

$$\dot{x} = x(\alpha - x) - xu\,.$$

Here, the control $u(t) \in [0,1]$ denotes the intensity of fishing activity. Given an initial population $x(0) = \bar{x}$, we want to maximize the payoff function

$$\int_0^T x(t)u(t) - \kappa u^2(t)\,dt\,,$$

which represents the total of fish caught mins the cost of fishing, where we assume $\alpha, \kappa, \bar{x} > 0$.

(i) Write Pontryagin's optimality conditions for this problem. (5 marks)

(ii) Give an explicit expression for the optimal control $u^*(t)$ depending on the state variable x, its adjoint λ, and model parameters. (5 marks)

(b) Consider the dynamics

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = u$$

where the control signal is limited by $|u| \leq 1$. One wants to minimize the criterion

$$\int_0^1 |u|dt + x_1(1)^2$$

for a given initial value.

(i) Write Pontryagin's optimality conditions for this problem. (5 marks)

(ii) Show that the necessary conditions for optimality imply that u changes finitely many times between $-1, 0$ and 1 . What is the maximum number of switches? (5 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2023

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. (ai) The function is expressed as

meth seen ⇓

$$f(\mathbf{x}) := \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x}\,,$$

with

$$\mathbf{A} := \begin{bmatrix} 1 & \frac{\alpha}{2} \\ \frac{\alpha}{2} & 1 \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} 1 \\ \beta \end{bmatrix},$$

for which $\nabla f = 2\mathbf{A}\mathbf{x} + \mathbf{b}$, and the existence of stationary points is linked to the existence of solutions to $\nabla f = 0 \Leftrightarrow 2\mathbf{A}\mathbf{x} = -\mathbf{b}$.

For the system to have a unique solution, we need $\det(\mathbf{A}) = 1 - \frac{\alpha^2}{4} \neq 0$, that is $\alpha \neq \pm 2$ (for any value of β). In this case, there exists a unique stationary point given by $\mathbf{x}^* = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$.

In the case $\alpha = \pm 2$, there exist infinite stationary points for $\beta = \pm 1$, respectively, lying along $2x + 2y = -1$. Otherwise there are no stationary points. 7, A

(aii) Given a quadratic function, if $\nabla^2 f = 2\mathbf{A}$ is positive semidefinite, the stationary points (assuming a solution exists) correspond to global minimisers. We have $tr(\mathbf{A}) = 2$ and we require $\det(\mathbf{A}) \geq 0$, which corresponds to $\alpha \in [-2, 2]$. For the extreme cases $\alpha = \pm 2$ we need $\beta = \pm 1$, respectively.

meth seen ⇓

5, A

(bi) The function $f(x, y) = x^2 - 2xy^2 + \frac{1}{2}y^4$ is not coercive. We can express it as

sim. seen ⇓

$$f(x, y) = x^2 - 2xy^2 + \frac{1}{2}y^4 = (x - y^2)^2 - \frac{1}{2}y^4\,,$$

and we can consider trajectories of the form (y^2, y) so that

$$\lim_{y\to\infty} f(y^2, y) = \lim_{y\to\infty} -\frac{1}{2}y^4 = -\infty\,.$$

4, B

(bii) Stationary points solve $\nabla f = 0$ or equivalently

sim. seen ⇓

$$\begin{aligned} x &= y^2 \\ 2y(-2x + y^2) &= 0\,, \end{aligned}$$

from where it follows that the only stationary point is given by $(0, 0)$. In this case the Hessian is given by

$$\nabla^2 f = \begin{bmatrix} 2 & -4y \\ -4y & -4x + 6y^2 \end{bmatrix}, \quad \nabla^2 f(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},$$

which is positive semi-definite. In this case, the stationary point can be either a local minimizer or a saddle point. Using the same trajectory as in part (bii) we can reach the origin with values below $f(0, 0) = 0$, and by taking trajectories $(x, 0)$ we can reach the origin with positive values. Hence, the origin is a saddle point. 4, B

2. (a) We have that

$$f'(x) = 2x^2 sign(x) + x\,.$$

If we apply gradient descent with $t^k = \frac{1}{k+1}$ we obtain

$$x_{k+1} = x_k\left(1 - \frac{2|x_k| + 1}{k+1}\right)$$

Using the hint, if we assume $x_{k+1} \geq k+1$ we have

$$x_{k+1} = \frac{x_k(k - 2x_k)}{k+1} \leq -(k+2)\,,$$

and similarly, if $x_k \leq -(k+1)$ we deduce

$$x_{k+1} = \frac{x_k(k + 2x_k)}{k+1} \geq k+2\,,$$

We conclude that

$$|x_k| \geq k+1 \Rightarrow |x_{k+1}| \geq k+2\,.$$

Since $|x_0| \geq 1$, by induction, we obtain that $|x_k| \geq k+1$, for all k, and the sequence diverges.

unseen ⇓

8, D

(bi) The function g is a quadratic function where

seen ⇓

$$\nabla^2 g = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \succ 0, \quad tr(\nabla^2 g) = 4\,, \quad \det(\nabla^2 g) = 3\,,$$

hence, convex. The function $f(t) = e^t$ is convex ($f'' > 0$ for all t) and non-decreasing, hence the composition is convex.

6, A

(bii) First we note that the feasible set is a convex set (intersection of half-planes). Minimising $-h(x,y)$ is equivalent to maximise $h(x,y)$ and maximisation of a convex function over a convex set is known to have a maximiser among the extreme points of the set. In this case the extreme points are given by $(0,0)$, $(0,1)$ and $(0.5, 0.5)$. It suffices to evaluate h in these three points:

seen/sim.seen ⇓

$$h(0,0) = 1$$
$$h(0,1) = e^2$$
$$h(0.5, 0.5) = e^{-0.25}\,,$$

from where we deduce that a minimiser of the original problem is $(0,1)$.

6, A

3. (a) The definition of stationarity indicates that for $\mathbf{x}^*$ to be a stationary point, it must satisfy

$$\nabla f(\mathbf{x}^*)^\top(\mathbf{x}-\mathbf{x}^*) \geq 0\,, \quad \forall \mathbf{x} \in C_{\mathbf{l},\mathbf{u}}\,.$$

sim. seen ⇓

We write the rhs equivalently as

$$\sum_{l_i<x_i^*<u_i}^{n} \frac{\partial f(\mathbf{x}^*)}{\partial x_i}(x_i - x_i^*) \;+\; \sum_{x_i^*=u_i}^{n} \frac{\partial f(\mathbf{x}^*)}{\partial x_i}(x_i - u_i^*) \;+\; \sum_{x_i^*=l_i}^{n} \frac{\partial f(\mathbf{x}^*)}{\partial x_i}(x_i - l_i^*)\,,$$

which we show to be non-negative as the first term in the sum vanishes as for every term $\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0$, the second sum non-negative as $\frac{\partial f(\mathbf{x}^*)}{\partial x_i} \leq 0$ and $(x_i - u_i^*) \leq 0$ since $\mathbf{x} \in C$, and similarly, for the last sum it holds $\frac{\partial f(\mathbf{x}^*)}{\partial x_i} \geq 0$ and $(x_i - l_i^*) \geq 0$.

8, B

(b) We observe that the cost is convex in $\mathbf{x}$, C is also convex, and hence KKT conditions are necessary and sufficient for this problem. We will show that a KKT point (i.e., a minimizer) satisfies the stationarity condition. The Lagrangian is given by

sim. seen ⇓

$$L(\mathbf{x},\lambda,\beta) = \|\mathbf{x}-\mathbf{b}\|_2^2 + \sum \lambda_i(l_i - x_i) + \sum \beta_i(x_i - u_i)\,,$$

from where it follows that the KKT system is given by

$$\begin{aligned} x_i - b_i - \lambda_i + \beta_i &= 0\,, \\ \lambda_i(l_i - x_i) &= 0\,, \\ \beta_i(x_i - u_i) &= 0\,, \\ \lambda_i, \beta_i &\geq 0\,, \qquad i = 1,\ldots,n\,. \end{aligned}$$

Each coordinate is studied independently, we proceed case by case:

i) $\lambda_i = \beta_i = 0$, in which case $x_i^* = b_i$, provided that $l_i < b_i < u_i$ for feasibility. In this case $\partial_{x_i} f(\mathbf{x}^*) = 2(x_i - b_i) = 0$.

ii) $\lambda_i > 0\,, \beta)i > 0$, which is unfeasible as the complementary slackness conditions cannot be satisfied simultaneously.

iii) $\lambda_i > 0\,, \beta_i = 0$. Here it follows that $x_i^* = l_i$ and hence $\lambda_i = l_i - b_i$, which is the solution when $l_i > b_i$. In this case $\partial_{x_i} f(\mathbf{x}^*) = 2(l_i - b_i) \geq 0$.

iv) $\lambda_i = 0$, $\beta_i > 0$. Similarly as in iii), it follows that $x_i^* = u_i$ and $\beta_i = b_i - u_i$, which is the solution when $b_i > u_i$. In this case $\partial_{x_i} f(\mathbf{x}^*) = 2(u_i - b_i) \leq 0$.

Note that depending on b_i, only one of the three feasible scenarios is possible.

8, C

(c) For a convex function such as $f = \|A\mathbf{x} - \mathbf{c}\|_2^2$ over a convex set C, we can use projected gradient descent, which requires, the computation of the orthogonal projection of the current value $\mathbf{x}^k$ onto C. However, this is precisely the routine that is given to us, and corresponds to the computation done in part b. Recalling that $\nabla f = 2A^\top(A\mathbf{x} - \mathbf{c})$, a general iteration would read

meth seen ⇓

$$\mathbf{x}^{k+1} = P_C[\mathbf{x}^k - 2t^k A^\top(A\mathbf{x}^k - \mathbf{c})]\,,$$

where P_C corresponds to the orthogonal projection subroutine with input $\mathbf{b} = \mathbf{x}^k$.

4, C

4. (ai) This is directly checked by noting that the origin is the unique feasible point. seen ⇓

3, A

(aii) The Lagrangian is given by: unseen ⇓

$$L(x, y, \lambda_1, \lambda_2) := x + \lambda_1(x^2 + (y-1)^2 - 1) + \lambda_2(x^2 + (y+1)^2 - 1)\,,$$

from where the KKT system is given by

$$\begin{aligned} 1 + 2\lambda_1 x + 2\lambda_2 x &= 0\,, \\ 2\lambda_1(y-1) + 2\lambda_2(y+1) &= 0\,, \\ \lambda_1(x^2 + (y-1)^2 - 1) &= 0\,, \\ \lambda_2(x^2 + (y+1)^2 - 1) &= 0\,, \\ \lambda_1, \lambda_2 &\geq 0\,. \end{aligned}$$

From where it is readily seen that the global minimizer $(0,0)$ does not satisfy the KKT conditions. Slater's condition is not satisfied, therefore KKT conditions are not necessary.

5, A

(bi) The set C is convex. As seen in lectures, C is an ellipsoid that is expressed as $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x}^\top Q\mathbf{x} \leq 1\}$, where $Q := diag(1,2,3) \succ 0$, hence convex. For an element $\mathbf{y}$, the orthogonal projection onto C is expressed as sim. seen ⇓

$$P_C(\mathbf{y}) := \underset{\mathbf{x}\in C}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{y}\|^2\,.$$

The fact that C is a convex set guarantees there exists a unique minimizer to the orthogonal projection problem.

4, B

(bii) This is a convex cost with convex constraint, Slater's condition is satisfied ($\mathbf{x} := 0 < 1$) , hence KKT conditions are necessary and sufficient. The Lagrangian is expressed as: sim. seen ⇓

$$L(\mathbf{x}, \lambda) := \|\mathbf{x} - \mathbf{y}\|^2 + \lambda(x_1^2 + 2x_2^2 + 3x_3^2 - 1)$$

and the KKT system reads

$$\begin{aligned} \mathbf{x} - \mathbf{y} + \lambda \begin{pmatrix} x_1 & 2x_2 & 3x_3 \end{pmatrix}^\top &= 0\,, \\ \lambda(x_1^2 + 2x_2^2 + 3x_3^2 - 1) &= 0\,, \\ \lambda &\geq 0\,. \end{aligned}$$

If $\mathbf{y} \in C$, then $\mathbf{x} = \mathbf{y}$ is the optimal solution to the orthogonal projection problem (and to the KKT system with $\lambda = 0$). We assume therefore $\lambda > 0$, which implies $x_1^2 + 2x_2^2 + 3x_3^2 = 1$. From the KKT system we also obtain:

$$\begin{aligned} x_1 &= \frac{1}{1+\lambda} y_1\,, \\ x_2 &= \frac{1}{1+2\lambda} y_2\,, \\ x_3 &= \frac{1}{1+3\lambda} y_3\,, \end{aligned}$$

leading into the nonlinear equation for λ

$$F(\lambda) := \left(\frac{1}{1+\lambda} y_1\right)^2 + 2\left(\frac{1}{1+2\lambda} y_2\right)^2 + 3\left(\frac{1}{1+3\lambda} y_3\right)^2 - 1 = 0\,.$$

Here, we note that $F(\lambda)$ is continuous for positive λ, $F(0) > 0$ (since$\mathbf{y}$ is outside C) and $\lim_{\lambda\to\infty} F(\lambda) = -1$, hence by the intermediate value theorem there exists a λ^* value for which $F(\lambda^*) = 0$, determining the unique orthogonal projection.

8, D

5. (ai) $\ast$ **Option 1:** We consider the problem $\boxed{\text{sim. seen} \Downarrow}$

$$\max \int_0^T x(t)u(t) - \kappa u^2(t)\ dt.$$

The Hamiltonian is given by:

$$H(x, \lambda, u) := xu - \kappa u^2 + \lambda(x(\alpha - x) - xu)$$

and PMP reads

$$\begin{aligned}
\dot{x} &= \frac{\partial H}{\partial \lambda} = x(\alpha - x) - xu, \quad x(0) = \bar{x}, \\
\dot{\lambda} &= -\frac{\partial H}{\partial x} = -(\alpha x - 2x - u)\lambda - u, \quad \lambda(T) = 0, \\
u^* &= \operatorname*{argmax}_{w \in [0,1]} H(x, \lambda, w)
\end{aligned}$$

$\ast$ **Option 2:** We consider the problem

$$\min \int_0^T \kappa u^2(t) - x(t)u(t)\ dt.$$

The Hamiltonian is given by:

$$H(x, \lambda, u) := \kappa u^2 - xu + \lambda(x(\alpha - x) - xu)$$

and PMP reads

$$\begin{aligned}
\dot{x} &= \frac{\partial H}{\partial \lambda} = x(\alpha - x) - xu, \quad x(0) = \bar{x}, \\
\dot{\lambda} &= -\frac{\partial H}{\partial x} = -(\alpha x - 2x - u)\lambda + u, \quad \lambda(T) = 0, \\
u^* &= \operatorname*{argmin}_{w \in [0,1]} H(x, \lambda, w)
\end{aligned}$$

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

(aii) $\ast$ **Option 1:** From the verification theorem in PMP, it follows that the optimal control is given by $\boxed{\text{meth seen} \Downarrow}$

$$\begin{aligned}
u^* &= \operatorname*{argmax}_{w \in [0,1]} H(x, \lambda, w) \\
&= \operatorname*{argmax}_{w \in [0,1]} \left\{ -\kappa w^2 + x(1 - \lambda)w + \lambda x(\alpha - x) \right\} \\
&= \operatorname*{argmax}_{w \in [0,1]} \left\{ -\kappa w^2 + x(1 - \lambda)w \right\},
\end{aligned}$$

which corresponds to a quadratic maximisation with a constrained variable. Using stationary conditions and defining the switching variable

$$w := \frac{x(1 - \lambda)}{2\kappa},$$

we obtain

$$u^*(x, \lambda, \kappa) = u^*(w) = \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \le w \le 1, \\ 1 & \text{otherwise.} \end{cases}$$

* **Option 2:** From the verification theorem in PMP, it follows that the optimal control is given by

$$\begin{aligned} u^* &= \underset{w\in[0,1]}{\text{argmin}}\ H(x,\lambda,w) \\ &= \underset{w\in[0,1]}{\text{argmin}}\ \left\{\kappa w^2 - (1+\lambda)xw + \lambda x(\alpha - x)\right\} \\ &= \underset{w\in[0,1]}{\text{argmin}}\ \left\{\kappa w^2 - (1+\lambda)xw\right\}, \end{aligned}$$

which corresponds to a quadratic minimisation with a constrained variable. Using stationary conditions and defining the switching variable

$$w := \frac{x(1+\lambda)}{2\kappa},$$

we obtain

$$u^*(x,\lambda,\kappa) = u^*(w) = \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \le w \le 1, \\ 1 & \text{otherwise.} \end{cases}$$

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

(bi) The Hamiltonian for the problem is $\boxed{\text{sim. seen} \Downarrow}$

$$H(u(t),x(t),\lambda(t)) = |u(t)| + \lambda_1(t)x_2(t) + \lambda_2(t)u(t)$$

and the terminal cost

$$\phi(x) = x_1$$

The adjoint equations will then be

$$\begin{aligned} \dot\lambda_1(t) &= -\frac{\partial H}{\partial x_1} = 0 \\ \dot\lambda_2(t) &= -\frac{\partial H}{\partial x_2} = \lambda_1(t)\,, \end{aligned}$$

with terminal conditions

$$\lambda_1(1) = \frac{\partial\psi}{\partial x_1}(x^*(1)) = 2x_1^*(1)\,, \quad \lambda_2(1) = \frac{\partial\psi}{\partial x_2}(x^*(1)) = 0\,.$$

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

(bii) This gives $\lambda_1(t) = C_1$ and $\lambda_2(t) = -C_1 t + C_2$. Furthermore, with the boundary constraints we get $\boxed{\text{meth seen} \Downarrow}$

$$\lambda_1(1) = \frac{\partial\psi}{\partial x_1}(x^*(1)) = 2x_1^*(1) = C_1\,, \quad \lambda_2(1) = \frac{\partial\psi}{\partial x_2}(x^*(1)) = 0\,.$$

The control signal is chosen as

$$u = \underset{|w|\le1}{\text{argmin}} H(x,w,\lambda) = \underset{|w|\le1}{\text{argmin}}\left\{|w| + \lambda_2 w\right\} \begin{cases} -1 & \lambda_2 > 1 \\ 0 & |\lambda_2| \le 1 \\ 1 & \lambda_2 < -1 \end{cases}$$

Since λ_2 is a linear function which ends at $\lambda_2(1) = 0, u$ can change value at maximum one time, either from - 1 to 0 or from 1 to 0. $\boxed{5,\ \mathrm{M}}$

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

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.		
ExamModuleCode	**QuestionNumber**	**Comments for Students**
MATH60005/70005	1	1a. In general, the students did well with this question. It students lost marks, it was because they either did not consider, or considered incorrectly the cases of \alpha = \pm 2.1b. Again in general, the students did well with this question. Students that did lose marks did so by assuming or incorrectly determining that the function was coercive. Students tended to lose marks also by not showing directly that the origin is a saddle point.
MATH60005/70005	2	2a. In general, the students handled this one fairly well. Students that lost marks tended to do so by being a bit careless and not showing directly that f is convex, or not obtaining the correct extremal points. Some students considered the full KKT system and marks were awarded in this case if the analysis was correct and all relevant cases were considered.2b. The results of this question were mixed. Most students were able to set up the interations correctly, but struggled use the hint to show that the interation would diverge if the initial guess is \geq 1.
MATH60005/70005	3	3 a) Most of the students stated the stationarity condition, followed by a direct verification that the proposed characterization satisfies the condition. Some students made an attempt in the wrong direction, by trying a proof by contradiction that is wrong in this case. b) Here, it was expected that students would use convexity and KKT conditions to show that the minimizer which solves KKT indeed verifies the stationarity conditions. Instead, many students proceed to directly find a minimizer by inspection as they correctly identified the problem as an orthogonal projection into a box. Students who did these calculations and justified their reasoning, either by computation or mentioning the link to orthogonal projection received full marks. c) Most of the students made a partial attempt to this question. Many correctly identified projected gradient descent as the algorithm to implement but made mistakes either in indicating what was the project, and what is the inner argument to be projected, receiving partial marks. Other students attempted to solve (P) by a change of variables but failed to apply the same change of variables to the set C, which leads to a set different from a box, so the projection from previous items cannot be used.
MATH60005/70005	4	4a) The majority of the students correctly identified (0,0) as the only feasible point, hence the unique global minimizer. Some students attempted to solve via coercivity/convexity, but this is not relevant for this problem. Many students were able to identify that (0,0) does not satisfy the KKT conditions due to Slater's condition not being fulfilled. Some students claimed (0,0) satisfies KKT, which is not correct. b) Overall, convexity of C was correctly identified by students arguing via lectures and/or convexity of ellipsoids. Incomplete explanations just indicating positive definiteness of a matrix involved in the definition of C did not receive full marks. Orthogonal projection was ok, but not everybody mentioned that due to convexity of C, the orthogonal projection is unique. For the computation of the orthogonal projection onto C, many students correctly attempted via KKT but did not discuss the solvability of the nonlinear equation for the multiplier lambda, receiving partial marks.
MATH70005	5	5a) Overall, correctly executed by many students, except for minor errors of sign, or inconsistencies when using argmin or argmax formulations. For aii), a few students were not able to perform explicitly the Hamiltonian minimization. b) As in a) a few sign errors, and not identifying that t=1 is the terminal time. In bii), some students did not compute lambda_2 so they were unable to determine a maximum number of switches. Some students did not realize that the terminal condition lambda_2(1)=0 implies at maximum one switch instead of two.