Temporary front page

BSc, MSc and MSci EXAMINATIONS (MATHEMATICS)

May – June 2022

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

Optimisation Mock Exam

Date: Wednesday, 11th May 2021

Time: 09:00-11:00

Time Allowed: 2 Hours for MATH96 paper; 2.5 Hours for MATH97 papers

This paper has *4 Questions (MATH96 version); 5 Questions (MATH97 versions)*.

Statistical tables will not be provided.

- Credit will be given for all questions attempted.
- Each question carries equal weight.

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

$$f(x_1, x_2) := x_2^4 - 2x_2^2 + 1 + \left(x_1^2 + x_2^2 - 1\right)^2$$

(i) Find all the stationary points of f. (5 marks)

(ii) Classify the stationary points found in i). (5 marks)

Answer i) The gradient of f is given by

$$\nabla f(\mathbf{x}) = 4 \begin{pmatrix} (x_1^2 + x_2^2 - 1)\, x_1 \\ (x_1^2 + x_2^2 - 1)\, x_2 + (x_2^2 - 1)\, x_2 \end{pmatrix}$$

The stationary points are those satisfying

$$\begin{aligned} \left(x_1^2 + x_2^2 - 1\right) x_1 &= 0 \\ \left(x_1^2 + x_2^2 - 1\right) x_2 + \left(x_2^2 - 1\right) x_2 &= 0\,. \end{aligned}$$

By the first equation, there are two cases: either $x_1 = 0$, and then by the second equation x_2 is equal to one of the values $0, 1, -1$; the second option is that $x_1^2 + x_2^2 = 1$, and then by the second equation we have that $x_2 = 0, \pm 1$ and hence x_1 is ± 1,0 respectively. Overall, there are 5 stationary points: $(0,0), (1,0), (-1,0), (0,1), (0,-1)$.

Rubric example: this is normally a type A question, and the marking would give 2 marks for computing the right nonlinear system $\nabla f(x) = 0$ and the remaining 3 marks for the stationary points.

Answer ii) The Hessian of the function is

$$\nabla^2 f(\mathbf{x}) = 4 \begin{pmatrix} 3x_1^2 + x_2^2 - 1 & 2x_1 x_2 \\ 2x_1 x_2 & x_1^2 + 6x_2^2 - 2 \end{pmatrix}.$$

Since

$$\nabla^2 f(0,0) = 4 \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \prec 0$$

it follows that (0,0) is a strict local maximum point. Since $f(x_1, 0) = (x_1^2 - 1)^2 + 1 \to \infty$ as $x_1 \to \infty$, the function is not bounded above and thus (0,0) is not a global maximum point. Also,

$$\nabla^2 f(1,0) = \nabla^2 f(-1,0) = 4 \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$$

which is an indefinite matrix and hence (1,0) and (-1,0) are saddle points. Finally,

$$\nabla^2 f(0,1) = \nabla^2 f(0,-1) = 4\begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \succeq 0\,.$$

The fact that the Hessian matrices of f at (0,1) and (0,-1) are positive semidefinite is not enough in order to conclude that these are local minimum points; they might be saddle points. However, in this case it is not difficult to see that (0,1) and (0,-1) are in fact global minimum points since $f(0,1) = f(0,-1) = 0$, and the function is lower bounded by zero. Note that since there are two global minimum points, they are nonstrict global minima, but they actually are strict local minimum points since each has a neighborhood in which it is the unique minimizer.

In summary: $(0,0)$ is a strict local maximum point, $(1,0)$ and $(-1,0)$ are saddle points, $(0,1)$ $(0,-1)$ are strict local minimum points (or non-strict global minima).

Rubric example: this is normally a type A question, 1 mark for correctly classifying each point, 1/2 marks for incomplete but correct characterization.

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex as well as concave function. Show that f is an affine function, that is, there exist $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$. (10 marks)

Answer. Since f is concave and convex, it follows immediately that $f(\lambda x+(1-\lambda)y) = \lambda f(x)+ (1-\lambda)f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. It is straightforward (though a bit tedious) to use this to check that the function $g(x) = f(x) - f(0)$ is linear, hence $g(x) = a^T x$, so letting $f(0) = b$, $f(x) = a^T x + b$ is affine.

Rubric example: this is normally a type D question since it includes a proof -although simple- 4 marks for $f(\lambda x + (1-\lambda)y) = \lambda f(x) +$ $(1-\lambda)f(y)$ 2 marks for working with $g(x) = f(x) - f(0)$, last 4 marks for a proof of linearity of $g(x)$.

(Total: 20 marks)

2. (a) Let $f(x_1, x_2)$ be a twice-differentiable convex function in $\mathbb{R}^2$ such that $f(0,0) = f(1,0) = f(0,1) = 0$. What do you know about:

(i) $f\left(\frac{1}{2}, \frac{1}{2}\right)$? (5 marks)

(ii) $a = \frac{\partial^2 f}{\partial x_1^2}$, $b = \frac{\partial^2 f}{\partial x_2^2}$, and $c = \frac{\partial^2 f}{\partial x_1 \partial x_2}$? (5 marks)

Answer

a.i) Noting that $\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}(1,0) + \frac{1}{2}(0,1)$, using the definition of convexity we have

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = f\left(\frac{1}{2}(1,0) + \frac{1}{2}(0,1)\right) \leq \frac{1}{2}f(0,1) + \frac{1}{2}f(1,0) \leq 0\,.$$

Rubric example: this is normally a type B question, 2 marks for expressing $(0.5, 0.5)$ as a convex combination of the points, 3 marks for working out the inequality upon the convexity definition.

a.ii) For a twice-differentiable smooth convex function we have that its Hessian is positive semidefinite, which applying determinant and trace criterion means $a + b \geq 0$ and $ab - c^2 \geq 0$.

Rubric example: this is normally a type A question, 2 for convex = Hessian positive semidefinite, 3 marks for applying a positive definiteness criterion and characterize in terms of a, b, c.

(b) Consider the function

$$g(x_1, x_2, x_3) = 59x_1^2 + 52x_2^2 + 17x_3^2 + 80x_1x_2 - 24x_1x_3 + 8x_2x_3 + 27x_1 - 84x_2 + 20x_3 \,.$$

(i) Is $g(\mathbf{x})$ convex? (5 marks)

(ii) Solve

$$\begin{aligned} \max \quad & g(x_1, x_2, x_3) \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

(5 marks)

Answer

i The function g is a quadratic function associated matrix

$$\mathbf{A} = \begin{bmatrix} 59 & 40 & -12 \\ 40 & 52 & 4 \\ -12 & 4 & 17 \end{bmatrix}$$

which is positive definite (diagonally dominant with positive entries in the diagonal). Therefore g is convex.

Rubric example: this is normally a type A question, 2 marks for expressing as quadratic form, 3 marks for checking positive definiteness.

ii The function g is non-constant, continuous and convex over the convex set of the constraints, which corresponds to the unit simplex Δ_3. Therefore, there exists at least one maximimizer of g over Δ_3 that is an extreme point of Δ_3. The extreme points of Δ_3 are given by the canonical basis in $\mathbb{R}^3$. It suffices to evaluate $g(1, 0, 0) = 86$, $g(0, 1, 0) = -32$, and $g(0, 0, 1) = 37$ to conclude that the maximizer is $g(1, 0, 0) = 86$.

Rubric example: this is normally a type A question, 3 marks for stating theorem for maximization of convex functions and commenting on convexity of cost and constraint, 2 marks for evaluating extreme points and determining maximizer.

(Total: 20 marks)

3. Consider the minimization problem for $\mathbf{x} \in \mathbb{R}^n$

$$\min_{\mathbf{x}\in C} \|\mathbf{x}-\mathbf{y}\|^2 ,$$

where $C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$, with $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{y} \in \mathbb{R}^n$. Assume that the rows of $\mathbf{A}$ are linearly independent.

(i) Determine the KKT conditions for this problem. Are these sufficient?. (6 marks)

(ii) Find the optimal solution of the problem using the KKT system. (8 marks)

(iii) Given the problem

$$\min_{(x_1,x_2)\in C} x_1^2 + 2x_2^2 - 3x_1 ,$$

where $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1\}$, write the explicit gradient descent iteration for this problem with a constant stepsize $t = 1$. You can help yourself using the result in part ii) (6 marks)

Answer

i) This is a convex optimization problem, so the KKT conditions are necessary and sufficient. The Lagrangian function is

$$L(\mathbf{x}, \lambda) = \|\mathbf{x}-\mathbf{y}\|^2 + 2\lambda^\top(\mathbf{A}\mathbf{x}-\mathbf{b}) = \|\mathbf{x}\|^2 - 2\left(\mathbf{y} - \mathbf{A}^\top\lambda\right)^\top \mathbf{x} - 2\lambda^\top\mathbf{b} + \|\mathbf{y}\|^2, \quad \lambda \in \mathbb{R}^m$$

Therefore, the KKT conditions are

$$2\mathbf{x} - 2\left(\mathbf{y} - \mathbf{A}^\top\lambda\right) = 0,$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Rubric example: this is normally a type B question, 2 marks noting sufficiency based on convexity, 2 marks for Lagrangian, 2 marks for KKT system.

ii) The first equation can be written as

$$\mathbf{x} = \mathbf{y} - \mathbf{A}^\top\lambda$$

Substituting this expression for $\mathbf{x}$ in the second equation yields the equation

$$\mathbf{A}\left(\mathbf{y} - \mathbf{A}^\top\lambda\right) = \mathbf{b}$$

which is the same as

$$\mathbf{A}\mathbf{A}^\top\lambda = \mathbf{A}\mathbf{y} - \mathbf{b}$$

Thus,

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

where here we used the fact that $\mathbf{A}\mathbf{A}^\top$ is nonsingular since the rows of $\mathbf{A}$ are linearly independent. Using the latter expression for λ, we obtain that the optimal solution is

$$\mathbf{x}^* = \mathbf{y} - \mathbf{A}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{b})$$

Rubric example: this is normally a type C question, 4 marks for finding optimal λ, 2 marks for noting that $A^\top A$ is invertible, 2 marks for expressing optimal x.

iii) Solving the problem in ii) corresponds to computing the orthogonal projection of a vector $\mathbf{y}$ onto the affine space C. Applying this directly to the problem in c) with $\mathbf{A} = [1\ 1]$ and $\mathbf{b} = 1$, the orthogonal projection of a vector $\mathbf{y}$ onto C is given by

$$P_C(\mathbf{y}) = \mathbf{y} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}(y_1 + y_2 - 1)\,.$$

The projected gradient descent with $t = 1$ reads

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

where

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 - 3 \\ 4x_2 \end{bmatrix}\,.$$

Rubric example: this is normally a type B question, 2 marks for identifying projection operator, 2 points applying previous results correctly, 2 marks for the gradient iteration.

(Total: 20 marks)

4. Consider the problem

$$\begin{aligned} \min \quad & x_1^2 + 0.5x_2^2 + x_1x_2 - 2x_1 - 3x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 1 \end{aligned}$$

(i) Solve this problem using KKT conditions. Are these sufficient? (10 marks)

(ii) Find the solution of the dual problem. What is the duality gap? (10 marks)

Answer In the revision session.

(Total: 20 marks)

5. **Mastery question.** The dynamics

$$\dot{x}(t) = -x(t) + u(t) \qquad |u| \leq 1$$

are to be controlled so that $x(1) = 0$ while minimizing the cost

$$J = \int_0^1 |u(t)|\, dt\,.$$

Show that the control

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

satisfies Pontryagin's necessary optimality conditions for some $x(0)$.

(20 marks)

Answer The Hamiltonian is

$$H(x, u, \lambda) = |u| + \lambda(u - x) \quad \text{and} \quad \Phi(x) = 0, \quad \Psi(x) = x$$

The multipliers λ shall fulfill

$$\begin{aligned} \dot{\lambda}(t) &= -H_x\left(x^*, u^*, \lambda\right) = \lambda(t) \\ \lambda(1) &= \Phi_x\left(x^*(1)\right) + \mu \Psi_x\left(x^*(1)\right) = \mu \end{aligned}$$

This means that $\lambda(t) = \mu e^{t-1}$, this means positive and increasing or negative and decresing. We will minimize the Hamiltonian with respect to u for each time instance. Is is equivalent to minimizing

$$H_1(u, \lambda) = |u| + \lambda u = \underbrace{(\operatorname{sign}(u) + \lambda)}_{\sigma} u$$

If $\sigma > 0$ we want to choose u as the smallest feasible negative value and if $\sigma < 0$ we want to choose u as the largest feasible positive value. Consequently, the following cases minimize the Hamiltonian: Positive λ

$$\begin{cases} u = -1 \text{ and } \lambda > 1 \\ u = 0 \text{ and } 0 \leq \lambda \leq 1 \end{cases}$$

Negative λ

$$\begin{cases} u = 1 \text{ and } \lambda < -1 \\ u = 0 \text{ and } -1 \leq \lambda \leq 0 \end{cases}$$

Our control candidate contains the control signals $0, -1$ which requires a positive λ. Thus, we have to find a μ such that λ starts with a positive value less than 1 and passes 1 at the time 0.5. This results in

$$1 = \mu e^{-\frac{1}{2}} \Rightarrow \mu = e^{\frac{1}{2}}$$

This μ fulfills the requirements and the optimality conditions are satisfied for the sudgested control.

As usual, with PMP you'll get marks for: stating the Hamiltonian, expressing PMP adjoint equations from here, working the proper initial/terminal conditions, and working out the structure that arises for the optimal control when minimizing the Hamiltonian. Further marks for carrying on with the requested computations based on this.

(Total: 20 marks)