Math31042 Fractal Geometry: Summer 2019 Exam Solutions

March 5, 2019

Students should be able to

- (ILO1) define the Hausdorff measure, Hausdorff dimension and box dimension of a set, explaining why any limits involved in the definition are well defined. State and prove lemmas from the course involving these concepts,
- (ILO 2) give upper bounds for the Hausdorff and box dimensions of sets by constructing suitable sequences of covers,
- (ILO 3) give lower bounds for the Hausdorff and box dimension of sets using the mass distribution principle, Lipschitz maps, and separation arguments,
- (ILO 4) define an attractor of an iterated function system and prove that such attractors exist and are unique,
- (ILO 5) state and prove the formula for the Hausdorff dimension of self-similar sets.

A1) a) A δ-cover of E is a collection $\{U_j\}_{j\in J}$ of sets $U_j \subset X$ satisfying that $E \subset \bigcup_{j\in J} U_j$ and that each U_j has $|U_j| \leq \delta$.

$$\mathcal{H}^s_\delta(E) := \inf\{\sum_{j\in J} |U_j|^s : \{U_j\}_{j\in J} \text{ is a } \delta\text{-cover of } E\}.$$

[3 marks, bookwork]

b)

$$\mathcal{H}^s(E) := \lim_{\delta\to 0} \mathcal{H}^s_\delta(E).$$

For any $\delta' < \delta$ we have that any δ'-cover of E is also a δ-cover of E. Thus

$$\mathcal{H}^s_{\delta'}(E) \geq \mathcal{H}^s_\delta(E).$$

Since $\mathcal{H}^s_\delta(E)$ is monotone in δ we conclude that the limit as $\delta \to 0$ exists. [4 marks, bookwork]

c) An outer measure μ on X is a function $\mu : \mathcal{P}(X) \to [0, \infty]$ satisfying

1. $\mu(\emptyset) = 0$
2. $A \subset B \subset X$ implies that $\mu(A) \leq \mu(B)$
3. If $A_j \in \mathcal{P}(X)$ for all $j \in \mathbb{N}$ then

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

If $A \subset B$ then any δ-cover of B is also a δ-cover of A. Hence

$$\mathcal{H}^s_\delta(A) \leq \mathcal{H}^s_\delta(B).$$

Taking limits as $\delta \to 0$ gives $\mathcal{H}^s(A) \leq \mathcal{H}^s(B)$. [5 marks, bookwork]

d) Define

$$\dim_H(E) = \inf\{s : \mathcal{H}^s(E) = 0\} = \sup\{t : \mathcal{H}^t(E) = \infty\}.$$

Let $A \subset B$. Then since $\mathcal{H}^s$ is an outer measure, we have that

$$\mathcal{H}^s(B) = 0 \Rightarrow \mathcal{H}^s(A) = 0.$$

Hence

$$\inf\{s : \mathcal{H}^s(A) = 0\} \leq \inf\{s : \mathcal{H}^s(B) = 0\}$$

and so $\dim_H(A) \leq \dim_H(B)$. [bookwork, 5 marks] **Q1 all ILO 1**

A2 a) An outer measure is a mass-distribution if it has bounded support and if

$$0 < \mu(X) < \infty.$$

[Alternative statement for last part, $0 < \mu(supp(\mu)) < \infty$.]
[3 marks, bookwork]

c) **Statement:** Let μ be a mass distribution on E, and suppose that for some s there are numbers $c > 0$ and $\epsilon > 0$ such that

$$\mu(u) \leq c|U|^s$$

for all sets $U \subset E$ with $|U| < \epsilon$. Then

$$\mathcal{H}^s(E) \geq \frac{\mu(E)}{c}$$

and hence $s \leq \dim_H(E)$.

Proof: If $\{U_j\}_{j\in J}$ is any δ-cover of E where $\delta < \epsilon$ then

$$\mu(E) \leq \mu\left(\bigcup_{j\in J} U_j\right) \leq \sum_{j\in J} \mu(U_j) \leq c\sum_{j\in J} |U_j|^s.$$

Rearranging gives

$$\sum_{j\in J} |U_j|^s \geq \frac{\mu(E)}{c} > 0$$

for any δ-cover of E, and so $\mathcal{H}^s(E) > 0$. Hence $\dim_H(E) \geq s$. [6 marks, bookwork]

d) Consider one dimensional Lebesgue measure λ_1 on $[0, 1]$. If $|U| < \delta$ then

$$\lambda_1(U) = |U| \leq \delta^1.$$

Thus $1 \leq \dim_H([0, 1])$ by the mass distribution principle.
[4 marks, example sheet] **Q2 ILOs 1 and 2**

A3 a) $\mathcal{S}(X)$ is the set of non-empty compact subsets of X. The Hausdorff metric d_H on $\mathcal{S}(X)$ is given by

$$d_H(A,B) := \inf\{\delta > 0 : A \subset B_\delta \text{ and } B \subset A_\delta\}$$

where for a set $C \in \mathcal{S}(X)$ the set $C_\delta := \{x \in X : \exists c \in C \text{ with } d(x,c) \leq \delta\}$.

$$d_H(\{0\},[1,2]) = 2.$$

[5 marks, bookwork/simple example]

A3 b) An iterated function system $\{\phi_1, \cdots \phi_N\}$ satisfies the open set condition if there is a non-empty open set $Y \subset X$ such that $\phi_i(Y) \subset Y$ for each $i \in \{1, \cdots, N\}$ and such that

$$\phi_i(Y) \cap \phi_j(Y) = \phi$$

whenever $i \neq j$.

The Sierpinski gasket is the self-similar set associated to the three contractions $\phi_1, \phi_2, \phi_3 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\phi_1(x,y) = \left(\frac{x}{2}, \frac{y}{2}\right), \phi_2(x,y) = \left(\frac{x+1}{2}, \frac{y}{2}\right), \phi_3(x,y) = \left(\frac{x}{2} + \frac{1}{4}, \frac{y}{2} + \frac{\sqrt{3}}{4}\right)$$

which satisfies the open set condition. [4 marks, bookwork]

A3 c) For the iterated function system $\Phi = \{\phi_1, \cdots \phi_k\}$, let the similarities ϕ_i have contraction ratio $c_i < 1$ then the similarity dimension is the unique s such that

$$\sum_{i=1}^{k} c_i^s = 1.$$

If each c_i takes the same value c then the above equation reduces to

$$kc^s = 1$$

and hence

$$s = \frac{\log(k)}{-\log(c)}.$$

[4 marks, bookwork]**Q3 Pre-ILO 3**

A4) Let $\mathcal{U} = \{U_j\}_{j\in J}$ be a δ-cover or E. Then the collection

$$\mathcal{U} + 1 := \{U_j + 1\}_{j\in J}$$

is a δ-cover of $E + 1$. Then

$$\begin{aligned} \mathcal{H}^s_\delta(E+1) &\leq \sum_{j\in J} |U_j + 1|^s \\ &= \sum_{j\in J} |U_j|^s. \end{aligned}$$

Taking the infimum over all δ covers gives $\mathcal{H}^s_\delta(E+1) \leq \mathcal{H}^s_\delta(E)$, and reversing the role of E and $E+1$ gives the reverse inequality. Hence

$$\mathcal{H}^s_\delta(E+1) = \mathcal{H}^s_\delta(E)$$

for all $\delta > 0$, letting $\delta \to 0$ gives $\mathcal{H}^s(E) = \mathcal{H}^s(E+1)$, and hence $\mathcal{H}^s(E) = 0 \iff \mathcal{H}^s(E+1) = 0$. Taking the infimum of the s for which the previous equations hold gives

$$\dim_H(E+1) = \dim_H(E)$$

as required. [5 marks, variant on exercise sheet] **Q4 ILOs 2 and 3**

B5, a) Standard sketch

[3 marks, easy variant on standard example]

B5 b) Let $\delta > 0$ and choose $k \in \mathbb{N}$ such that

$$\sqrt{2}.3^{-k} \leq \delta < \sqrt{2}.3^{-(k-1)}.$$

Then E can be covered by the 3^k sets

$$\phi_{a_1} \circ \cdots \phi_{a_k}([0,1]^2)$$

where each $a_i \in \{1,2,3\}$. These sets have diameter $\sqrt{2}.3^{-k} \leq \delta$, so this gives a δ-cover of E. Hence $\mathcal{N}_\delta(E) \leq 3^k$. Then

$$\begin{aligned}\frac{\log(\mathcal{N}_\delta(E))}{-\log(\delta)} &\leq \frac{3^k}{-\log(3^{-(k-1)})} \\ &= \frac{k}{k-1}\frac{\log 3}{\log 3}.\end{aligned}$$

Taking limsup as $\delta \to 0, k \to \infty$ gives $\overline{\dim}_B(E) \leq 1$.

[5 marks, easy variant on standard examples]

B5 c) Consider the map $\pi_1 : \mathbb{R}^2 \to \mathbb{R}$ given by $\pi_1(x,y) = x$. This is a Lipschitz map since

$$d(\pi_1(x_1,x_2), \pi_1(y_1,y_2)) = |x_1 - y_1| \leq \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2} = d((x_1,x_2),(y_1,y_2)).$$

Hence by a theorem from the course,

$$\dim_H(\pi_1(E)) \leq \dim_H(E).$$

But $\pi_1(E) = [0,1]$ which has Hausdorff dimension 1, and so

$$\dim_H(E) \geq 1$$

as required.

[6 marks, easy variant on standard example (Cantor dust)]. **Q5 ILOs 2 and 3**

B6 a) $f : (X, d) \to (X', d')$ is α-Hölder continuous with constant K if for all $x, y \in X$ with $d(x, y) \leq 1$,

$$d'(f(x), f(y)) \leq K d(x, y)^\alpha$$

Let $\{U_j\}_{j \in J}$ be a δ-cover of E for $\delta \leq 1$. Then $|f(U_j)| < K\delta^\alpha$ and so $\{f(U_j)\}_{j \in J}$ is a $K\delta^\alpha$ cover of $f(E)$.

Then

$$\mathcal{H}^{\frac{s}{\alpha}}_{K\delta^\alpha}(f(E)) \leq \sum_{j \in J} (K|U_j|^\alpha)^{\frac{s}{\alpha}} = K^{\frac{s}{\alpha}} \sum_{j \in J} |U_j|^s$$

This holds for each δ-cover of E, so taking the infimum gives

$$\mathcal{H}^{\frac{s}{\alpha}}_{K\delta^\alpha}(f(E)) \leq K^{\frac{s}{\alpha}} (\mathcal{H}^s_\delta(E))$$

Taking the limit as $\delta \to 0$ gives the required statement. [bookwork 6 marks]

B6 b) If $f : (X, d) \to (X', d')$ is α-Hölder and $E \subset X$, $0 < \alpha \leq 1$, then

$$\dim_H(f(E)) \leq \frac{1}{\alpha} \dim_H(E)$$

[bookwork 2 marks]

B6 c) Consider the function $f : F \to [0, 1]$ given by

$$f(\sum_{i=1}^{\infty} a_i 3^{-i}) = \sum_{i=1}^{\infty} \frac{a_i}{2} 2^{-i}.$$

Since each $x \in F$ can be represented as $\sum_{i=1}^{\infty} a_i 3^{-i}$ with $a_i \in \{0, 2\}$, this maps F surjectively onto $[0, 1]$. If $x, y \in F$ have $3^{-k} \leq |x - y| < 3^{-(k-1)}$ then the sequences $\underline{a}$ corresponding to x and $\underline{b}$ corresponding to y agree to the first $k - 1$ places. So

$$|f(x) - f(y)| \leq \sum_{n=k}^{\infty} \frac{1}{2^n} = 2^{-(k-1)}.$$

So

$$|f(x) - f(y)| \leq (3|x - y|)^\alpha$$

where $\alpha = \frac{\log 2}{\log 3}$, and so f is α-Hölder. Since $\dim_H([0,1]) = 1$, we have from the previous theorem that

$$1 = \dim_H([0,1]) \leq \frac{1}{\alpha}\dim_H(F)$$

and so $\dim_H(F) \geq \alpha = \frac{\log 2}{\log 3}$. [8 marks, exercise sheet] **Q6 ILO 3**

B7 a) Let $\Phi = \{\phi_1, \cdots \phi_k\}$ be an iterated function systems made of contractive maps on X. $\Psi : X \to X$ is given by

$$\Psi(A) = \bigcup_{i=1}^{n} \phi_i(A).$$

Hutchinson's theorem states that there is a unique set $E \in \mathcal{S}(X)$, that is a unique compact non-empty subset of X, for which $\Psi(E) = E$. Furthermore, for any $A \in \mathcal{S}(X)$, $\Psi^n(A) \to E$ where convergence is in the Hausdorff metric.

For each $i \in \{1, \cdots r\}$ there exists $c_i < 1$ such that $\phi_i : X \to X$ is a contraction with contraction constant c_i. Letting $c := \max\{c_1, \cdots c_r\}$ gives that

$$d(\phi_i(x), \phi_i(y)) < cd(x, y)$$

for all $i \in \{1, \cdots, r\}$ and $x, y \in X$.

Let $A, B \in \mathcal{S}(X)$. Now if $d_H(A, B) = \delta$ then for all $a \in A$ there exists $b \in B$ such that $d(a, b) \leq \delta$.

Then for all j and for all $\phi_j(a) \in \phi_j(A)$ there exists $\phi_j(b) \in \phi_j(B)$ such that

$$d(\phi_j(a), \phi_j(b)) \leq cd(a, b) \leq c\delta.$$

Thus $\phi_j(A) \subset (\phi_j(B))_{c\delta}$, and since this holds for all $j \in \{1, \cdots, r\}$ we have

$$\Psi(A) \subset (\Psi(B))_{c\delta}.$$

Reversing the role of A and B in the above calculation gives $\Psi(B) \subset (\Psi(A))_{c\delta}$ and hence that

$$d_H(\Psi(A), \Psi(B)) \leq cd_H(A, B).$$

Using the contraction mapping theorem on the complete metric space $(\mathcal{S}(X), d_H)$ gives that there is a unique 'point' $E \in \mathcal{S}(X)$ with $\Psi(E) = E$ and such that, for any $A \in \mathcal{S}(X)$,

$$d_H(\Psi^n(A), E) \to 0$$

as $n \to \infty$.

[8 marks, bookwork]

B7 b) Let $\phi_1(x) = \frac{x}{3}$, $\phi_2(x) = \frac{x+1}{2}$, and let s be such that

$$\left(\frac{1}{3}\right)^s + \left(\frac{1}{2}\right)^s = 1.$$

Let $c_1 = \frac{1}{3}$, $c_2 = \frac{1}{2}$. Define a measure μ on E by letting

$$\mu(\phi_{a_1} \circ \cdots \phi_{a_n}(E)) = c_{a_1}^s \cdots c_{a_n}^s,$$

and extend this to an outer measure on all subsets A of E by letting

$$\mu(A) = \inf\{\sum_{j\in J} \mu(\phi_{a_1^j} \circ \cdots \phi_{a_{n_j}^j}(E)) : A \subset \bigcup_{j \in J} \phi_{a_1^j} \circ \cdots \phi_{a_{n_j}^j}(E)\}.$$

Let $U \subset E$ have $|U| < 1$. For $\underline{x} \in \{1, \cdots, r\}^{\mathbb{N}}$ let

$$k_{\underline{x}} := \min\{k \in \mathbb{N} : c_{x_1} \cdots c_{x_{k_{\underline{x}}}} \leq |U|\}$$

This gives us a finite cover of E

$$\mathcal{U} = \{\phi_{x_1} \circ \cdots \phi_{x_{k_{\underline{x}}}}([0,1]) : \underline{x} \in \{1, \cdots, r\}^{\mathbb{N}}\}$$

made of disjoint sets of roughly the same diameter (which is roughly equal to $|U|$).

Note that

$$\mu(\phi_{x_1} \circ \cdots \phi_{x_{k_{\underline{x}}}}([0,1])) = c_{x_1}^s \cdots c_{x_{k_{\underline{x}}}}^s \leq (|U|)^s$$

Now since $c_{x_1} \cdots c_{x_{k_{\underline{x}}-1}} > |U|$ we see that the disjoint intervals $\phi_{x_1} \circ \cdots \phi_{x_{k_{\underline{x}}}}((0,1))$ have diameter at least $\frac{|U|}{3}$, and so U can intersect at most 3 such intervals. In fact, U is covered by these 3 intervals (since E is covered by the collection $\mathcal{U}$, and $U \subset E$. Then

$$\begin{aligned} \mu(U) &\leq \sum_{i=1}^{3} \mu(\phi_{x_1^i} \circ \cdots \phi_{x_{k_{\underline{x}^i}}^i}([0,1]) \\ &\leq 3|U|^s. \end{aligned}$$

Hence by the mass distribution principle we conclude that $\mathcal{H}^s(E) > 0$ and $\dim_H(E) \geq s$ as required. [8 marks, special case of bookwork]

ILOs 4 and 5

1.

(a) Define what is meant by a *topology* on a set X.

(b) Define what is meant by the *usual* topology on a subset $X \subset \mathbb{R}^n$.

(c) Define what is meant by saying that a function $f\colon X \to Y$ between topological spaces is *continuous*. Define what is meant by saying that f is a *homeomorphism*.

(d) Consider the set $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\}$. Show that τ defines a topology on X. Find all continuous maps from $X \to \{1, 2\}$ with the usual topology on $\{1, 2\}$.

[15 marks]

Solution

(a) Given a set X, a *topology* on X is a collection τ of subsets of X with the following properties:

(i) $\emptyset \in \tau$, $X \in \tau$;

(ii) the intersection of any two subsets in τ is in τ:

$$U_1, U_2 \in \tau \Rightarrow U_1 \cap U_2 \in \tau;$$

(iii) the union of any collection of subsets in τ is in τ:

$$U_\lambda \in \tau \text{ for all } \lambda \in \Lambda \Rightarrow \bigcup_{\lambda \in \Lambda} U_\lambda \in \tau.$$

[6 marks, bookwork]

(b) The usual topology on $X \subset \mathbb{R}^n$ is the subspace topology, i.e $U \subset X$ is open if and only if for every $x \in X$ there is an $\epsilon > 0$, such that $B_\epsilon(x) \cap X \subset U$. [2 marks, bookwork]

(c) $f\colon X \to Y$ is *continuous* if

$$V \text{ is open in } Y \Rightarrow f^{-1}(V) \text{ is open in } X$$

[1 marks, bookwork]

A *homeomorphism* is a continuous bijection with continuous inverse.

[2 marks, bookwork]

(d)
- $\emptyset$, $X \in \tau$,
- the only non-trivial intersection is $\{a, b\} \cap \{c\} = \emptyset \in \tau$,

- the only non-trivial union is $\{a, b\} \cup \{c\} = X \in \tau$.

First note, that every subset of $\{1, 2\}$ is open in the usual topology (hence the usual topology coincides with the discrete topology).

A map $\psi\colon X \to \{0, 1\}$ is a continuous, if and only if $\psi^{-1}(\{0\}) \in \tau$ and $\psi^{-1}(\{1\}) \in \tau$ (since $\psi^{-1}(\emptyset) = \emptyset$ and $\psi^{-1}(\{0, 1\}) = X$ are automatically open).

Assume $\psi^{-1}(0) = X$ this implies $\psi(x) = 0$ for all $x \in X$. Hence, $\psi = \mathrm{const}_0$ similarly if $\psi^{-1}(0) = \emptyset$ this implies $\psi(x) \neq 0$, hence, $\psi(x) \neq 1$ for all $x \in X$ and $\psi = \mathrm{const}_1$. Now assume $\psi^{-1}(0) = \{a, b\}$ it follows, that $\psi(a) = \psi(b) = 0$ and $\psi(c) = 1$. Similarly for $\psi^{-1}(0) = \{c\}$ we conclude $\psi(a) = \psi(b) = 1$ and $\psi(c) = 0$. Hence, we have the four continuous functions

(i) $a \mapsto 1, b \mapsto 1, c \mapsto 1$

(ii) $a \mapsto 0, b \mapsto 0, c \mapsto 0$

(iii) $a \mapsto 1, b \mapsto 1, c \mapsto 0$

(iv) $a \mapsto 0, b \mapsto 0, c \mapsto 1$

Alternatively, one can check all the $8 = 2^3$ maps $X \to \{0, 1\}$ for their continuity by applying the definition.

[4 marks, similar to question set]

[Total: 15 marks]

2.

(a) Define what is meant by saying that a topological space X is *path-connected.*

(b) What is meant by saying that path-connectedness is a *topological property*?

(c) Define what is meant by the set $\pi_0(X)$ of *path-components* of a topological space X.

(d) Prove that a continuous map of topological spaces $f\colon X \to Y$ induces a map $f_*\colon \pi_0(X) \to \pi_0(Y)$ between the sets of path-components, taking care to prove that your function is well-defined. Prove that if f is a homeomorphism then f_* is a bijection.

(e) Use (d) to show that path-connectedness is a topological property.

(f) Consider the topological space (X, τ) from Question 1. (d), what are the path-components of this space? Justify your answer.

[15 marks]

Solution

(a) A *path* from x_0 to x_1 in X is a continuous function $\sigma\colon [0, 1] \to X$ with $\sigma(0) = x_0$ and $\sigma(1) = x_1$.

X is said to be *path-connected* if, for each pair of points x_0, $x_1 \in X$, there is a path in X from x_0 to x_1.

[3 marks, bookwork]

(b) Saying that path-connectedness is a *topological property* means that, if $X \cong Y$ are homeomorphic topological spaces, then X is path connected if and only if Y is path-connected.

[1 marks, bookwork]

(c) Define an equivalence relation on X by $x \sim x'$ if and only if there is a path in X from x to x'. Then the path-components of X are the equivalence classes and $\pi_0(X)$ is the set of all equivalence classes.

[2 marks, bookwork]

(d) Suppose that $f\colon X \to Y$ is a continuous map. Then this induces a function $f\colon \pi_0(X) \to \pi_0(Y)$ by $f([x]) = [f(x)]$. This is well-defined because $[x] = [x']$ implies that $x \sim x'$ so that there is a path $\sigma\colon [0,1] \to X$ in X from x to x'. Then $f \circ \sigma\colon [0,1] \to Y$ is a path in Y from $f(x)$ to $f(x')$ and so $[f(x)] = [f(x')]$.

[3 marks, bookwork]

If f is a homeomorphism then f_* is a bijection since the inverse $g = f^{-1}\colon Y \to X$ induces a function $g_* : \pi_0(Y) \to \pi_0(X)$ inverse to f_* since $g_*(f_*([x])) = [g(f(x))] = [x]$ and $f_*(g_*([y])) = [y]$. [2 marks, bookwork]

(e) Being path-connected is by definition equivalent to $\#\pi_0(X) = 1$. Hence, this follows directly from (d). [1 marks, bookwork]

(f) The path components are $\{a, b\}$ and $\{c\}$. To see this one has to show, that there is a path in X from a to b, but that there is no path from either a to c nor b to c. For the first statements we can simply state a path, e.g. the one given by

$$\sigma(t) = \begin{cases} a & t = 0, \\ b & t \in (0,1] \end{cases}$$

This is continuous, as $\sigma^{-1}(X) = [0,1] = \sigma^{-1}(a,b) = [0,1]$ and $\sigma^{-1}(\emptyset) = \sigma^{-1}(\{c\}) = \emptyset$, which are both open in $[0,1]$.

Assume $\sigma\colon [0,1] \to X$ is path from a to c. Consider the function $\psi\colon X \to \{1,2\}$ with $\psi(a) = \psi(b) = 0$ and $\psi(c) = 1$. This is continuous by 1. (d). Now $\psi \circ \sigma\colon [0,1] \to \{1,2\} \subset \mathbb{R}$ is surjective and continuous, which is in contradiction with the intermediate value theorem.

Similarly on shows, that there is no path from b to c. [3 marks, similar to question set]

[Total: 15 marks]

3.

(a) Suppose that $q\colon X \to Y$ is a surjection from a topological space X to a set Y . Define the *quotient topology* on Y determined by q. State the *universal property* of the quotient topology.

(b) Suppose that $f\colon X \to Z$ is a continuous surjection from a compact topological space X to a Hausdorff topological space Z . Define an equivalence relation $\sim$ on X so that f induces a bijection $F\colon X/{\sim} \to Z$ from the identification space $X/{\sim}$ of this equivalence relation to Z. Prove that F is a homeomorphism. [*State clearly any general results which you use.*]

(c) Prove that the quotient space $(S^1 \times [0,1])/(S^1 \times \{1\})$ is homeomorphic to the closed unit disc D^2 (where S^1 and D^2 have the usual topology).

[15 marks]

Solution

(a) Given a topological space (X, τ) and a surjection $q\colon X \to Y$ the quotient topology on Y is given by

$$\{V \subset Y \mid q^{-1}(V) \in \tau\}.$$

The *universal property* of the quotient topology is: $f\colon Y \to Z$ to a topological space Z is continuous if and only if the composition $f \circ q\colon X \to Z$ is continuous.

[4 marks, bookwork]

(b) Given a continuous surjection $f\colon X \to Z$, define an equivalence relation on X by $x \sim x' \Leftrightarrow f(x) = f(x')$. Then we may define $F\colon X/{\sim} \to Z$ by $F([x]) = f(x)$. Since $[x] = [x'] \Leftrightarrow x \sim x' \Leftrightarrow f(x) = f(x')$ (by the definition of the equivalence relation), the function F is well-defined. Since $F([x]) = F([x']) \Leftrightarrow f(x) = f(x') \Leftrightarrow x \sim x'$ (by the definition of the equivalence relation) it follows that $[x] = [x']$ and F is injective. Since f is a surjection, $y = f(x)$ for some $x \in X$ and so $y = F([x])$. Hence F is a surjection. This shows that $F\colon X/{\sim} \to Z$ is a bijection. The map $F : X/{\sim} \to Z$ is continuous by the universal property since $F \circ q = f$ which is given as continuous, where $q\colon X \to X/{\sim}$ is the quotient map given by $q(x) = [x]$.

The space $X/{\sim} = q(X)$ is compact since it is the continuous image of a compact set. Hence F is a homeomorphism since it is a continuous bijection from a compact space to a Hausdorff space.

[7 marks, bookwork]

(c) Define $f\colon S^1 \times [0,1] \to D^2$ by $f(x,t) = (1-t)x$. This is a continuous surjection. Furthermore $f(x,t) = f(x',t')$ if and only if $(x,t) = (x',t')$ or $t = t' = 1$, i.e. $(x,t) = (x',t')$, or (x,t) and $(x',t') \in S^1 \times \{1\}$. Furthermore, $S^1 \times [0,1]$ is compact (a product of closed bounded subsets of Euclidean spaces) and D^2 is Hausdorff (a subspace of a Euclidean space). Thus the general result in (b) gives the required homeomophism $F\colon (S^1 \times [0,1])/(S^1 \times \{1\} \to D^2$.

[4 marks, similar to question set]

[Total: 15 marks]

4.

(a) Define what is meant by saying that a topological space is *Hausdorff*.

(b) Suppose that X and Y are topological spaces. Define the *product topology* on the Cartesian product $X \times Y$. [*It is not necessary to prove that this is a topology.*]

(c) Consider the Sierpinski topology $\tau = \{\emptyset, \{a\}, \{a,b\}\}$ on $X = \{a,b\}$. Write down the product topology on $X \times X$.

(d) Prove that if X and Y are Hausdorff spaces, then the product space $X \times Y$ is Hausdorff, as well.

(e) Prove that, if a topological space X is Hausdorff, then all singleton subsets $\{x\}$ of X (where $x \in X$) are closed. Give an example to show that the converse of this statement is false; that is, if a topological space has all singleton subsets closed, it is not necessarily Hausdorff.

[15 marks]

Solution

(a) The topological space X is *Hausdorff* if, for each distinct pair of points x, $y \in X$, there exist open sets U and V in X such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

[2 marks, bookwork]

(b) The product topology on $X \times Y$ has a basis

$$\{U \times V \mid U \text{ open in } X, V \text{ open in } Y\},$$

i.e. the open sets consist of all unions of such sets.

[2 marks, bookwork]

(c) The sets $U \times V$ with $U, V \in \tau$ are the following

$$\emptyset, X \times X, \{a\} \times X = \{(a,a),(a,b)\}, X \times \{a\} = \{(a,a),(b,a)\}$$

There is only one non-trivial union to consider: $\{(a,a),(a,b)\} \cup \{(a,a),(b,a)\} = \{(a,a),(a,b),(b,a)\}$. Adding it to the sets stated above we obtain

$$\{\emptyset, X \times X, \{(a,a),(a,b)\}, \{(a,a),(b,a)\}, \{(a,a),(a,b),(b,a)\}\}.$$

[3 marks, similar to question set]

(d) Suppose X and Y are Hausdorff spaces and $(x_1, y_1) \neq (x_2, y_2) \in X \times Y$. Then $x_1 \neq x_2$ or $y_1 \neq y_2$. If $x_1 \neq x_2$ then there are disjoint open subsets $U_1, U_2 \subset X$ such that $x_1 \in U_1$ and $x_2 \in U_2$. Then $U_1 \times Y$, $U_2 \times Y$ are disjoint open subsets of $X \times Y$ as required. There is a similar argument if $y_1 \neq y_2$. Hence $X \times Y$ is Hausdorff. [3 marks, bookwork]

(e) For each point $x \in X \setminus \{a\}$, since X is Hausdorff, there are disjoint open subsets U_x and V_x of X such that $a \in U_x$ and $x \in V_x$. Since the subsets are disjoint, $a \notin V_x$ and so $V_x \subset X \setminus \{a\}$. Hence $X \setminus \{a\} = \bigcup_{x \in X \setminus \{a\}} V_x$ and so, being a union of open subsets, is open in X. Hence $\{a\}$ is closed in X. [4 marks, question set]

The set $\mathbb{R}$ with the cofinite topology as all finite sets closed and so, in particular, singleton sets are closed but it is not Hausdorff. [1 mark, standard example]

[Total: 15 marks]

5.

(a) Define what is meant by a *compact subset* of a topological space.

(b) Prove by using only the definition of compactness, that $\{\frac{1}{n} \mid n \in \mathbb{N}\} \subset \mathbb{R}$ with the usual topology is not compact.

(c) Prove that every compact subset of a Hausdorff topological space is closed.

(d) Give an example to show that the Hausdorff condition in (c) cannot be omitted.

[15 marks]

Solution

(a) $K \subset X$ is compact if each cover of K by open subsets of X has a finite subcover.

[3 marks, bookwork]

(b) Consider the subsets $U_m = \{\frac{1}{n} \mid n \leqslant m\}$ for $m \in \mathbb{N}$. These are open, since $U_m = X \cap (\frac{1}{m+1}, 2)$. Then $X = \bigcup_{m\in\mathbb{N}} U_m$. Assume $\mathcal{F} \subset \{U_m \mid m \in \mathbb{N}\}$ is a finite subcover. Then there exists an $N \in \mathbb{N}$ such that $U_m \notin \mathcal{F}$ for $m \geqslant N$. Hence, $\frac{1}{N} \notin U$ for $U \in \mathcal{F}$ and consequently $\frac{1}{N} \notin \bigcup_{U\in\mathcal{F}} U$. [5 marks, bookwork]

(c) Suppose that K is a compact subset of a Hausdorff space X. To prove that K is closed we prove that $X \setminus K$ is open, and we prove this by proving that it is a union of open subsets. Let $x \in X \setminus K$. Then, by the Hausdorff condition, for each $a \in K$ there are open subsets U_a, V_a of X such that $a \in U_a$, $x \in V_a$ and $U_a \cap V_a = \emptyset$.

Then $\{U_a \mid a \in A\}$ is an open cover for K. Hence, since K is compact, there is a finite subcover $\{U_{a_i} \mid 1 \leqslant i \leqslant n\}$ for K. So $K \subset \bigcup_{i=1}^n U_{a_i}$.

Put $V_x = \bigcap_{i=1}^n V_{a_i}$. Then V_x is a finite intersection of open sets and so is open and, since $x \in V_{a_i}$ for all i, $x \in V_x$. Furthermore, for $1 \leqslant i \leqslant n$, $V_x \cap U_{a_i} \subset V_{a_i} \cap U_{a_i} = \emptyset$ and so $V_x \cap U_{a_i} = \emptyset$. Hence $V_x \cap \bigcup_{i=1}^n U_{a_i} = \emptyset$ and so $X_x \cap K = \emptyset$ or, equivalently, $V_x \subset X \setminus K$.

Thus $X \setminus K = \bigcup_{x\in X\setminus K} V_x$ is a union of open sets and so is open. Hence K is closed.

[5 marks, bookwork]

(d) Consider $X = [-1, 1]/\sim$, where $s \sim \pm s$ when $|s| < 1$. Then $U^+ = X \setminus \{[-1]\} \cong [0, 1]$ (via $q|_{[0,1]}$) is compact. But, $q^{-1}(U^+) = (-1, 1]$, which is not closed in $[-1, 1]$. Hence, U^+ is not closed.

There are also simple finite examples. [2 marks, bookwork]

[Total: 15 marks]

6.

(a) Suppose that X_1 is a subspace of a topological space X. Define what is meant by saying that X_1 is a *retract* of X.

(b) Show that $S^1 = \{x \in \mathbb{R}^2 \mid |x| = 1\}$ is a retract of the punctured plane $\mathbb{R} \setminus \{0\}$.

(c) Explain how a continuous function of topological spaces $f\colon X \to Y$ induces a homomorphism of fundamental groups $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ for $x_0 \in X$. You should indicate why f_* is a well-defined homomorphism.

(d) Use the functorial properties of the fundamental group to prove that, if X_1 is a retract of X, then, for any $x_0 \in X_1$, the homomorphism induced by the inclusion map

$$i_*\colon \pi_1(X_1, x_0) \to \pi_1(X, x_0)$$

is injective.

(e) Hence prove that S^1 is not a retract of the closed disc D^2 .

[*You may quote any calculation of fundamental groups that you need, without proof.*]

[15 marks]

Solution

(a) $X_1 \subset X$ is a retract of X when there is a continuous map $r\colon X \to X_1$, such that $r(x) = x$ for $x \in X_1$. [3 marks, bookwork]

(b) By the functorial properties we have

$$r_* \circ i_* = (r \circ i)_* = (\mathrm{id}_{X_1})_* = \mathrm{id}_{\pi_1(X_1,x_0)}\colon \pi_1(X_1, x_0) \to \pi_1(X_1, x_0).$$

Since the composition of r_* and i_* is bijective r_* must be surjective and i_* must be injective.

[4 marks, bookwork]

(c) The function f_* is defined by $f_*([\sigma]) = [f \circ \sigma]$. It is well-defined since, if $[\sigma_0] = [\sigma_1]$ then $\sigma_0 \sim \sigma_1$ and so there exists a homotopy $H\colon \sigma_0 \sim \sigma_1$. Then $f \circ H\colon I^2 \to Y$ gives a homotopy $f \circ \sigma_0 \sim f \circ \sigma_1$ and so $[f \circ \sigma_0] = [f \circ \sigma_1]$.

To see that f_* is a homomorphism suppose that $[\sigma]$, $[\tau] \in \pi_1(X, x_0)$. Then

$$f_*([\sigma][\tau]) = f_*([\sigma * \tau]) = [f \circ (\sigma * \tau)]$$

and

$$f_*([\sigma])f_*([\tau]) = [f \circ \sigma][f \circ \tau] = [(f \circ \sigma) * (f \circ \tau)]$$

and by writing out the formulae we see that $f \circ (\sigma * \tau) = (f \circ \sigma) * (f \circ \tau)\colon I \to Y$. Hence, $f_*([\sigma][\tau]) = f_*([\sigma])f_*([\tau])$. [5 marks, bookwork]

(d) We have $\pi_1(S^1, x_0) = \mathbb{Z}$ and $\pi_1(D^2, x_0) = 1$, the trivial group. But there is not injective map $\mathbb{Z} \to \{1\}$. Hence, S^1 cannot be a retract of D^2. [3 marks, bookwork]

END OF EXAMINATION PAPER

Intended learning outcome Question	1	2	3	4	5	6
prove that certain topological spaces are homeomorphic by constructing a concrete homeomorphism						
define the notions of path-connectedness and path-components and apply them to distinguish subsets of Euclidean space up to topological equivalence						
decide whether a collection of subsets of a set determines a topology and whether a map between topological spaces is continuous						
define the subspace topology, the product topology and the quotient topology, prove their universal properties and apply them to construct continuous maps						
recognise whether or not a topological space is compact or Hausdorff and state the basic properties of compact and Hausdorff spaces and their proofs						
define the fundamental group, use it to distinguish topological spaces, apply the functorial properties to find obstructions for the existence of particular continuous maps						

Color scheme: light=less challenging, dark = more challenging. Note, that there is a great variation within each question.

Solutions

A1, (3+3+4)

a) Riemannian metric G on n-dimensional manifold M^n defines for every point $\mathbf{p} \in M$ the scalar product of tangent vectors in the tangent space $T_{\mathbf{p}}M$ smoothly depending on the point $\mathbf{p}$. It means that in every coordinate system $(x^1, \dots, x^n)$ a metric $G = g_{ik}dx^i dx^k$ is defined by a matrix valued function $g_{ik}(x)$ $(i = 1, \dots, n; k = 1, \dots n)$ such that for any two vectors $\mathbf{A} = A^i(x)\frac{\partial}{\partial x^i}, \mathbf{B} = B^i(x)\frac{\partial}{\partial x^i}$, tangent to the manifold M at the point $\mathbf{p}$ with coordinates $x = (x^1, x^2, \dots, x^n)$ $(\mathbf{A}, \mathbf{B} \in T_{\mathbf{p}}M)$ the scalar product is equal to:

$$\langle \mathbf{A}, \mathbf{B} \rangle_G \big|_{\mathbf{p}} = G(\mathbf{A}, \mathbf{B})\big|_{\mathbf{p}} = A^i(x) g_{ik}(x) B^k(x) \,,$$

where

1. $G(\mathbf{A}, \mathbf{B}) = G(\mathbf{B}, \mathbf{A})$, i.e. $g_{ik}(x) = g_{ki}(x)$ (symmetricity condition)
2. $G(\mathbf{A}, \mathbf{A}) > 0$ if $\mathbf{A} \neq \mathbf{0}$, i.e.
   $g_{ik}(x)u^i u^k \geq 0$, $g_{ik}(x)u^i u^k = 0$ iff $u^1 = \dots = u^n = 0$ (positive-definiteness)
3. $G(\mathbf{A}, \mathbf{B})\big|_{\mathbf{p}=x}$,i.e. $g_{ik}(x)$ are smooth functions. ■

(b) Symmetricity condition $g_{12} = g_{21}$ means that $B \equiv C$.

Consider vector fields $\mathbf{X} = \frac{\partial}{\partial u}$ and $\mathbf{Y} = \frac{\partial}{\partial v}$. Since these both vector fields do not vanish, positive-definitness condition says that $G(\mathbf{X}, \mathbf{X}) = A > 0$ and $G(\mathbf{Y}, \mathbf{Y}) = D > 0$. ■

(c) If $\det g = 0$ hence this 2×2 matrix has non-zero eigenvector $\mathbf{A}$ with eigenvalue 0: $\mathbf{A} = a\partial_u + b\partial_v$ such that $a \neq 0$ or $b \neq 0$ and $\begin{pmatrix} A & B \\ B & D \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0$. Thus $G(\mathbf{A}, \mathbf{A}) = 0$. This contradicts to positive-definitness condition.

Remark *Positive definitness implies that* $\det g > 0$. *This was discussed during tutorials, but it is not an easy question.*

A2, (2+(5=2+3)+3)

a) n-dimensional Riemannian manifold (M, G) is locally Euclidean Riemannian manifold, if in a vicinity of every point $\mathbf{p}$ there exist local coordinates $u^1, \dots, u^n$ such that Riemannian metric G in these coordinates has an appearance

$$G = du^i \delta_{ik} du^k = (du^1)^2 + \dots + (du^n)^2 \,.$$

(b) For this cylindrical surface the induced Riemannian metric is

$$G = (dx^2 + dy^2 + dz^2)_{x=3\cos\varphi, y=3\sin\varphi, z=h} = (-3\sin\varphi d\varphi)^2 + (3\cos\varphi d\varphi)^2 + dh^2 = 9d\varphi^2 + dh^2 \,. ■$$

In a vicinity of every point one can consider new local coordinates $u = h, v = 3\varphi$ then $du^2 + dv^2 = dh^2 + 9d\varphi^2$. We see that in local coordinates u, v metric is Euclidean, i.e. thhese coordinates are Euclidean.

c) One can consider new local coordinates $\begin{cases} \tilde{u} = u + a \\ \tilde{v} = v + b \end{cases}$, where a, b are arbitrary constants then evidently $(d\tilde{u})^2 + (d\tilde{v})^2 = du^2 + dv^2$.

Another solution One can consider new local Euclidean coordinates (which are 'rotated') $\begin{cases} \tilde{u} = u\cos\varphi - v\sin\varphi \\ \tilde{u} = u\sin\varphi + v\cos\varphi \end{cases}$. These are also Euclidean coordinates.

A3, (3+5+(7=2+5))

a) Affine connection on M is the *operation* ∇ which assigns to every vector field $\mathbf{X}$ a linear map $\nabla_{\mathbf{X}}$ on the space of vector fields: $\nabla_{\mathbf{X}}(\lambda\mathbf{Y} + \mu\mathbf{Z}) = \lambda\nabla_{\mathbf{X}}\mathbf{Y} + \mu\nabla_{\mathbf{X}}\mathbf{Z}$ ($\lambda, \mu \in \mathbf{R}$), which satisfies the following additional conditions:

1 For arbitrary (smooth) functions f, g on M

$$\nabla_{f\mathbf{X}+\mathbf{gY}}(\mathbf{Z}) = f\nabla_{\mathbf{X}}(\mathbf{Z}) + g\nabla_{\mathbf{Y}}(\mathbf{Z}) \qquad (C^\infty(M)\text{-linearity})$$

2 For arbitrary function f

$$\nabla_{\mathbf{X}}(f\mathbf{Y}) = (\nabla_{\mathbf{X}}f)\,\mathbf{Y} + f\nabla_{\mathbf{X}}(\mathbf{Y}) \qquad (\text{Leibnitz rule})$$

($\nabla_{\mathbf{X}}f$ is just usual derivative of a function f along vector field: $\nabla_{\mathbf{X}}f = \partial_{\mathbf{X}}f$.) ■

(b) Using these properties we have

$$e^{-F}\nabla_{\mathbf{X}}(e^F\mathbf{Y}) = e^{-F}(\partial_{\mathbf{X}}e^F)\mathbf{Y} + e^{-F}\cdot e^F\nabla_{\mathbf{X}}\mathbf{Y} = e^{-F}\cdot e^F\partial_{\mathbf{X}}F\mathbf{Y} + \nabla_{\mathbf{X}}\mathbf{Y} = \partial_{\mathbf{X}}F\mathbf{Y} + \nabla_{\mathbf{X}}\mathbf{Y}\,.$$

(c) Canonical connection on Euclidean space $\mathbf{E}^n$ is a connection such that its Christoffel symbols vanish in Cartesian coordinates.

We have that $\begin{cases} x = r\cos\varphi \\ y = r\sin\varphi \end{cases}$, $r = \sqrt{x^2 + y^2}$, hence

$$\Gamma^r_{rr} = \frac{\partial^2 x}{\partial r^2}\frac{\partial r}{\partial x} + \frac{\partial^2 y}{\partial r^2}\frac{\partial r}{\partial y} = 0\,, \text{since } \tfrac{\partial^2 x}{\partial r^2} = \tfrac{\partial^2 y}{\partial r^2} = 0$$

and

$$\Gamma^r_{\varphi\varphi} = \frac{\partial^2 x}{\partial\varphi^2}\frac{\partial r}{\partial x} + \frac{\partial^2 y}{\partial\varphi^2}\frac{\partial r}{\partial y} = -r\cos\varphi\frac{x}{r} - r\sin\varphi\frac{y}{r} = -r\cos^2\varphi - r\sin^2\varphi = -r \quad ■$$

A4, (4+7+4)

a) Let M be a surface in $\mathbf{E}^3$, K be Gaussian curvature, $||g_{\alpha\beta}||$ induced Riemannian metric, R_{ikmn}—Riemann curvature tensor of Levi-Civita connection, R—scalar curvature of this tensor. Then

$$\frac{R}{2} = \frac{R_{1212}}{\det g} = K\,.$$

b) We have for the saddle

$$\mathbf{r}_u = \begin{pmatrix} 1 \\ 0 \\ kv \end{pmatrix}, \quad \mathbf{r}_v = \begin{pmatrix} 0 \\ 1 \\ ku \end{pmatrix}, \mathbf{n}(u,v) = \frac{1}{\sqrt{1 + k^2u^2 + k^2v^2}}\begin{pmatrix} -kv \\ -ku \\ 1 \end{pmatrix},$$

here $\mathbf{n}(u,v)$ is normal unit vector field to the surface at the point $\mathbf{p}$: $(\mathbf{r}_u,\mathbf{n})=(\mathbf{r}_v,\mathbf{n})=0$ and $(\mathbf{n},\mathbf{n})=1$.

Calculate shape operator at the point $u=v=0$:

$$S\left(\mathbf{r}_u\right)=-\frac{\partial\mathbf{n}(u,v)}{\partial u}\Big|_{u=v=0}=\begin{pmatrix}0\\ k\\ 0\end{pmatrix}=k\mathbf{r}_v\,, S\left(\mathbf{r}_v\right)=-\frac{\partial\mathbf{n}(u,v)}{\partial v}\Big|_{u=v=0}=\begin{pmatrix}k\\ 0\\ 0\end{pmatrix}=k\mathbf{r}_u\,,$$

$S\big|_{u=v=0}=\begin{pmatrix}0 & k\\ k & 0\end{pmatrix}$, and curvature at this point is equal to $K=\det S=-k^2$.

Thus the scalar curvature at the point $\mathbf{p}$ is equal to $R=2K=-2k^2$.

c) Suppose there exist local coordinates in a vicinity of a point $\mathbf{p}$ such that $G=du^2+dv^2$. Then in these coordinates all components of Christoffel symbols vanish $\Rightarrow$ $R=0$. Contradiction.

B5 (6+4+10)

The total area of the plane with respect to the stereographic metric F^*G is equal to the total area of the sphere without point= total area of the sphere $=4\pi$ since F is isometry of $\mathbf{R}^2$ onto $S^2\mathbf{N}$.

Another solution One can calculate it straightforwardly by `brute force`:
area of $\mathbf{R}^2$ with respect to the metric $G_{\text{stereographic}}$=

$$=\int_{\mathbf{R}^2}\sqrt{\det g}dudv=\int_{\mathbf{R}^2}\sqrt{\left(\frac{4}{(1+u^2+v^2)^2}\right)^2}dudv=\int_{\mathbf{R}^2}\frac{4}{(1+u^2+v^2)^2}dudv=$$

$$\int_{\mathbf{R}^2}\frac{4}{(1+r^2)^2}rdrd\varphi=2\pi\int_0^\infty\frac{2dz}{(1+z)^2}=-4\pi\frac{1}{1+z}\Big|_0^\infty=4\pi,.$$

(here we used polar coordinates, and did substitution $z=r^2$) The second solution is longer but not terribly longer.

Using stereogrpahic coordinates $\begin{cases}x=\frac{2u}{u^2+v^2+1}\\ y=\frac{2v}{u^2+v^2+1}\\ z=\frac{u^2+v^2-1}{u^2+v^2+1}\end{cases}$, we have $z=x\Leftrightarrow\frac{u^2+v^2-1}{1+u^2+v^2}=\frac{2u}{1+u^2+v^2}$, i.e. $u^2+v^2-1=2u$, i.e. $(u-1)^2+v^2=2$. i.e. C' is the circle in $\mathbf{R}^2$ of the radius $\sqrt{2}$ with centre at the point $u=1,v=0$.

c) The interior of the circle C' is the isometric image (stereographic projection) of the interior of the circle C on the sphere, i.e. the area of the interior of the circle C' is equal to the area of the half-sphere $=2\pi$.

Remark We can try to perform straightforward solution as for the case above using a `brute force`. We come to:

$$\text{Area of the interior of } C'=\int_{(u-1)^2+v^2\leq 2}\frac{4}{(1+u^2+v^2)^2}dudv=$$

$$\int_0^\infty dr\left(\int_0^{2\pi}d\varphi\frac{4rd\varphi}{(1+r^2+2r\cos\varphi)}\right).$$

Here we did substitution: $\begin{cases} u = 1 + r\cos\varphi \\ v = r\sin\varphi \end{cases}$
I do not think that it is realistic to calculate quickly this integral, and show that it is equal to 2π ■

B6(6+8+6)

a) Let M be a Riemannian manifold with metric $G = g_{ik}dx^i dx^k$. Let ∇ be a symmetric connection on M, i.e. its Christophel symbols Γ^m_{ik} satisfies the condition: $\Gamma^m_{ik} = \Gamma^m_{ki}$. We say that symmetric connection ∇ is Levi Civita connection if it preserves scalar product, i.e. if for arbitrary vectors $\mathbf{Y}, \mathbf{Z}$ at an arbitrary point

$$\partial_{\mathbf{X}} < \mathbf{Y}, \mathbf{Z} > = < \nabla_{\mathbf{X}}(\mathbf{Y}), \mathbf{Z} > + < \mathbf{Y}, \nabla_{\mathbf{X}}(\mathbf{Z}) > \ . \qquad (B6,1)$$

Levi-Civita Theorem claims that on the Riemannian manifold (M, G) there exists uniquely defined Levi Civita connection. In local coordinates Christoffel symbols of this connection have the following appearance:

$$G = g_{ik}dx^i dx^k \Rightarrow \quad \Gamma^m_{ik}(x) = \frac{1}{2} g^{mn}(x)\left(\frac{\partial g_{in}(x)}{\partial x^k} + \frac{\partial g_{kn}(x)}{\partial x^i} - \frac{\partial g_{ik}(x)}{\partial x^n}\right). \qquad (B6,2)$$

b) Use the Levi-Civita formula above. The derivatives of Riemannian metric vanish at the point $u = v = 0$. Indeed we have that

$$G = \sigma(du^2 + dv^2)\,, ||g_{\alpha\beta}|| = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}, \text{where } \sigma = \frac{4R^2}{(1 + u^2 + v^2)}\,, \qquad (B6.3)$$

hence for example

$$\frac{\partial g_{uu}}{\partial v}\Big|_{u=v=0} = \frac{\partial \sigma}{\partial v}\Big|_{u=v=0} = \frac{-2 \cdot 4R^2 \cdot 2v}{(1 + u^2 + v^2)}\Big|_{u=v=0} = 0\,.$$

Hence according to formula (B6,2) all Christoffel symbols vanish at the point $u = v = 0$ in stereographic coordinates.

c) Denote by $\tilde\Gamma^u_{uu}, \Gamma^u_{uv}, \ldots, \tilde\Gamma^v_{vv}$ Christoffel symbols of the connection $\tilde\nabla$. Consider preservation of scalar products (B6,2) by the connection $\tilde\nabla$. We have that $\langle\partial_u, \partial_u\rangle = \langle\partial_v, \partial_v\rangle = \sigma$, $\langle\partial_u, \partial_v\rangle = 0$ hence due to equation (B6,1)

$$0 = \partial_u\langle\partial_u, \partial_v\rangle = \langle\tilde\nabla_u\partial_u, \partial_v\rangle + \langle\partial_u, \tilde\nabla_u\partial_v\rangle = \tilde\Gamma^v_{uu}\langle\partial_v, \partial_v\rangle + \tilde\Gamma^u_{uv}\langle\partial_u, \partial_u\rangle =$$

$$\tilde\Gamma^v_{uu}\sigma + \tilde\Gamma^u_{uv}\sigma \Rightarrow \tilde\Gamma^u_{uv} \equiv 0\,, \text{since}\, \tilde\Gamma^v_{uu} \equiv 0\,,$$

and on the other hand

$$\partial_v\sigma = \partial_v\langle\partial_u, \partial_u\rangle = 2\tilde\Gamma^u_{vu}\sigma \Rightarrow \tilde\Gamma^u_{vu} = \frac{1}{2}\partial_v \log\sigma \not\equiv 0\,.$$

i.e. connection $\tilde\nabla$ is not symmetric: $0 \equiv \tilde\Gamma^u_{uv} \neq \tilde\Gamma^u_{vu}$.

We see that connection ∇ is not symmetric, hence it does not coincide with symmetric Levi-Civita connection ∇.

The existence of the non-symmetric connection which preserves scalar metric does not contradict to Levi-Civita theorem, which says about uniqueness of *symmetirc connection* which preserves the scalar metric.

B7(6+4+10)

a)Let $\mathbf{x} = \mathbf{x}(t)$, $t_1 \leq t \leq t_2$ be a parameterisation of curve C beginning at the point $\mathbf{p}_1$ and ending at the point $\mathbf{p}_2$. such that $\mathbf{x}(t_1) = \mathbf{p}_1$ and $\mathbf{x}(t_2) = \mathbf{p}_2$. Let $\mathbf{X} = \mathbf{X}_1 \in T_{\mathbf{p}_1}M$ be an arbitrary tangent vector at the initial point $\mathbf{p}_1$ of the curve C. (The vector $\mathbf{X}$ is not necessarily tangent to the curve C.) We say that $\mathbf{X}(t)$, $t_1 \leq t \leq t_2$ is a parallel transport of the vector $\mathbf{X}_1 \in T_{\partial_1}M$ along the curve $C: \mathbf{x} = \mathbf{x}(t), t_1 \leq t \leq t_2$, and the vector field $\mathbf{X}_2 = \mathbf{X}(t_2)$ attached at the point $\mathbf{p}_2$ is the result of parallel transport of the vector $\mathbf{X}_1$ from the point $\mathbf{p}_1$ to the point $\mathbf{p}_2$ along the curve C if the following conditions obey

i) For an arbitrary t, $t_1 \leq t \leq t_2$, vector $\mathbf{X}(t)$, $(\mathbf{X}_1 = \mathbf{X}(t_1))$ is a tangent vector attached at the point $\mathbf{x}(t)$, i.e. $\mathbf{X}(t)$ is a vector tangent to the manifold M at the point $\mathbf{x}(t)$ of the curve C, $\mathbf{X}(t) \in T_{\mathbf{x}(t)}M$, and in particularly the vector $\mathbf{X}_2 = \mathbf{X}(t_2) \in T_{\mathbf{p}_2}M$

ii) The covariant derivative of $\mathbf{X}(t)$ along the curve C equals to zero: $\frac{\nabla \mathbf{X}(t)}{dt} = \nabla_{\mathbf{v}}\mathbf{X} = 0$.. If $X^m(t)$ are components of the vector field $\mathbf{X}(t)$ and $v^m(t)$ are components of the velocity vector $\mathbf{v}$ of the curve C, then the equation $\frac{\nabla \mathbf{X}}{dt} = \nabla_{\mathbf{v}}\mathbf{X} = 0$ has in the components the following appearance:

$$\frac{dX^i(t)}{dt} + v^k(t)\Gamma^i_{km}(x^i(t))X^m(t) \equiv 0\,, \qquad t_1 \leq t \leq t_2\,. \qquad (B7,1)$$

Here ∇ is Levi-Civita connection.

One can say that the result of parallel transport of the vector $\mathbf{X}_1 \in T_{\mathbf{p}_1}M$ along a curve C is a vector $\mathbf{X}_2 \in T_{\mathbf{p}_2}M$ which is obtained from the solution of differentiao equation (1) with intial condition $\mathbf{X}(t)\big|_{t=t_1}$: $\mathbf{X}_2 = \mathbf{X}(t)\big|_{t=t_2}$.

We define parallel transport using a parameterisation of the curve C, One can show that the parallel transport $\mathbf{X} \to \mathbf{X}'$ remains the same if we choose another parameterisation of the curve C starting at the point $\mathbf{p}_1$, i.e. if new parameterisation has the same orientation.

The differential equation (B7,1) defining parallel transport is a linear equation, and the operator $P_C: T_{p_1}M \ni \mathbf{X}_1 \to \mathbf{X}_2 \in T_{p_2}M$ is linear operator. The connection ∇ defining parallel transport is Levi-Civita connection, it preserves scalar product:

$$\frac{d}{dt}\langle \mathbf{X}(t), \mathbf{X}(t)\rangle_{\mathbf{x}(t)} = \partial_{\mathbf{v}}\langle \mathbf{X}(t), \mathbf{X}(t)\rangle_{\mathbf{x}(t)} = 2\langle \nabla_{\mathbf{v}}X(t), \mathbf{X}(t)\rangle_{\mathbf{x}(t)} = 0\,,$$

due to the equation $\nabla_{\mathbf{v}}dX = 0$. We see that linear operator P_C, operator of of parallel transport, preserves scalar product, i.e. it is an orthogonal operator.

c) The plane C is on the distance $h = \frac{R\sqrt{2}}{2}$. of the origin. Hence the area of the segment of the spere D such that $C = \partial D$ is equal to

$$S(D) = 2\pi R(R - h) = 2\pi R^2\left(1 - \frac{\sqrt{2}}{2}\right)$$

thus the angle of rotation of attached vectors under parallel transport is equal to

$$\angle\Phi = \int_D K d\sigma = \frac{1}{R^2} S(D) = \pi(2-\sqrt{2}) \,.$$

and

$$P_C \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \cos\pi(2-\sqrt{2}) & -\sin\pi(2-\sqrt{2}) \\ \sin\pi(2-\sqrt{2}) & \cos\pi(2-\sqrt{2}) \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \cdot \blacksquare$$

Comments

INTENDNED LEARNING OUTCOME

`On completion of this unit succesfull students will be able`

I *state the definition of a Riemannian manifold M and calculate the length of a curve, and area of a domain in M*

II *calculate the Riemannian metric on surfaces embedded in $\mathbf{E}^3$*

III *define a connection on a manifold, state the Levi-Civita theorem, and calculate the connection for surfaces of cylinder, sphere and cone in $\mathbf{E}^3$, and for Lobachevsky (hyperbolic) plane*

IV *state the properties of geodesics on a Riemannian manifold, and calculate the parallel transport of vectors along a geodesics for the sphere and cylindre in $\mathbf{E}^2$ and for Lobachevsky plane*

V *State the definition of the Riemann curvature tensor. Calculate the Riemann curvature tensor for some 2-dimensional Riemannian manifolds.*

ILO of questions:

Question A1	*ILO I*
Question A2	*ILO I+II*
Question A3	*ILO III*
Question A4	*ILO II+V*
Question B5	*ILO I+II*
Question B6	*ILO III*
Question B7	*ILO IV+V*

First four questions A1—A4 worth 10+10+15+15=50 points. These questions are essentially bookwork and these questions are easy. Just to mention that

*question c) in **A1** (4 points) is not difficult, however may cause a problem.*

*question c) in **A2** is also easy but unexpected.*

*question b) in **A4** is bookwork but to answer this question you need to perform not very easy calculations.*

Next three questions B1,B2,B3 worth 20 points (every question). Student need to answer two of these questions. Maximum for these questions is 40/90 These questions are essentially more difficult. Just to mention that

Question **B5***— the question is difficult and the last part of the question is unseen. . It is almost impossible to calculate the integral with* `brute force`*.*

The half of the points of the question **B6** *a) is not easy but it is bookwork; to gain another half of the points student needs to special efforts, the last part of the queston is unseen.*

B7 *Question is not unseen, however to come to the final answer student has to guess to use the right formula.*

In total

Bookwork—-10+7+12+13+6+6+8=62
Easy questions—-6+7+10+8=31
Difficult or unseen—-0+3+0+0+0+10+8+10=31

MATH32032 **SOLUTIONS+MARKING SCHEME**

SECTION A

Answer **ALL** questions in this section (40 marks in total)

Below are the questions and the model solutions. The paper tested the Intended Learning Outcomes (ILO) of the course; ILO numbers next to each question refer to the following list.

ILO1 Define and illustrate main concepts and prove fundamental theorems concerning error-correcting codes, given in the course.

ILO2 Calculate the parameters of given codes and their dual codes using standard matrix and polynomial operations.

ILO3 Encode and decode information by applying algorithms associated with well-known families of codes.

ILO4 Compare the error-detecting/correcting facilities of given codes for a given binary symmetric channel.

ILO5 Design simple linear or cyclic codes with required properties.

ILO6 Solve mathematical problems involving error-correcting codes by linking them to concepts from elementary number theory, combinatorics, linear algebra and elementary calculus.

A1. (a) $\boxed{\text{ILO1, basic}}$ Define what is meant by:

Answer. [*bookwork — 5*]

— a *word* $\underline{x}$ of length n in a finite alphabet F;
[**Answer:** an element of $F^n = F \times \cdots \times F$ (*or equivalent*) $\boxed{1\text{m}}$]

— a *code* C of length n in a finite alphabet F;
[**Answer:** a non-empty subset of F^n $\boxed{1\text{m}}$]

— the *Hamming distance* $d(\underline{x}, \underline{y})$ between two words $\underline{x}$, $\underline{y}$ of length n;
[**Answer:** the number of positions where the two words differ $\boxed{1\text{m}}$]

— a *nearest neighbour* of a word $\underline{x}$ in a code C.
[**Answer:** a codeword $\underline{c} \in C$ such that $d(\underline{x}, \underline{c}) = \min\{d(\underline{x}, \underline{z}) : \underline{z} \in C\}$. $\boxed{2\text{m}}$]

(b) $\boxed{\text{ILO1, ILO3, basic}}$ Let C be the binary repetition code of length 6. List the codewords of C. Write down words $\underline{x}, \underline{y} \in \mathbb{F}_2^6$ such that $\underline{x}$ has exactly one nearest neighbour in C, and $\underline{y}$ has more than one nearest neighbour in C.

Answer. [*bookwork + examples similar to classwork — 5*]

$C = \{000000, 111111\}$. $\boxed{2\text{m}}$

A possible example is $\underline{x} = 000000$ and $\underline{y} = 000111$. $\boxed{3\text{m}}$

Comment. In fact, all the possible examples are as follows: $\underline{x}$ is a vector of weight other than 3, and $\underline{y}$ is a vector of weight 3 in $\mathbb{F}_2^6$. This is because vectors of weight 0, 1 or 2 have 000000 as their unique nearest neighbour in C, since their Hamming distance to 000000 is 0, 1 or 2 and their distance to 111111 is 6, 5 or 4. Similarly, vectors of weight ≥ 4 have unique nearest neighbour 111111 in C. Vectors of weight 3 are equidistant from 000000 and 111111 hence have two nearest neighbours in C.

[10 marks]

A2.

(a) $\boxed{\text{ILO1, ILO3, basic}}$ Explain what is meant by a *standard array* for a linear $[n, k, d]_q$-code C. State the number of rows and the number of columns in a standard array for C. Explain how to decode a received vector using the standard array.

Answer. [*bookwork — 5*]

A standard array is a table with $\#C = q^k$ columns and q^{n-k} rows. Each vector in $\mathbb{F}_q^n$ appears exactly once as an entry in the table. Each row is a coset of C, and the leftmost entry in the row is a coset leader. The top row is the trivial coset (i.e., C itself). Each entry is the sum of the leftmost entry in its row and the top entry in its column. $\boxed{3\text{m}}$

To decode a received vector $\underline{y} \in \mathbb{F}_q^n$, look up $\underline{y}$ in the standard array and return the codevector at the top of the column containing $\underline{y}$. (*Equivalently*, subtract from $\underline{y}$ the leftmost entry of the row containing $\underline{y}$.) $\boxed{2\text{m}}$

You are now given that $D = \{0000,\ 0111,\ \underline{v},\ 1110\}$ is a binary linear code.

(b) $\boxed{\text{ILO4, medium difficulty}}$ Prove that the vector $\underline{v}$ is 1001. Assuming that a codevector of D is sent via $\mathrm{BSC}(p)$, show that, with probability $p^2 - p^4$, the received vector will contain an error which is not detected.

Answer. [*standard, similar to classwork — 5*]

A linear code is closed under addition, hence D must contain $0111 + 1110 = 1001$, and this must be the vector $\underline{v}$. $\boxed{2\text{m}}$

The probability in question, $P_{\text{undetect}}(D)$, is given by the formula $\sum_{i=1}^{n} A_i(1-p)^{n-i}p^i$ where A_i is the number of codevectors of weight i. $\boxed{1\text{m}}$

In our case $n = 4$, $A_1 = 0$, $A_2 = 1$ and $A_3 = 2$ so $P_{\text{undetect}}(D)$ is $(1-p)^2p^2 + 2(1-p)p^3$ $\boxed{1\text{m}}$ which expands as $(1-p)p^2(1-p+2p) = (1-p)p^2(1+p) = p^2 - p^4$. $\boxed{1\text{m}}$

(c) $\boxed{\text{ILO3, basic}}$ Construct a standard array for D. Use the standard array you have constructed to give an example where one bit error occurs in a codevector $\underline{c}$ of D and is not corrected by the standard array decoder.

Answer. [*standard — 7*]

A possible standard array is $\begin{matrix} 0000 & 0111 & 1001 & 1110 \\ 0001 & 0110 & 1000 & 1111 \\ 0010 & 0101 & 1011 & 1100 \\ 0100 & 0011 & 1101 & 1010 \end{matrix}$. $\boxed{5\text{m}}$

Detailed steps (*not required in the exam*):

- The top row is the code, starting from the zero codevector but otherwise in any order.
- Next, consider vectors of weight 1. Pick any of them, say 0001, and make it the coset leader of the second row. Construct the second row by adding 0001 to the vectors in the top row.
- There are still vectors of weight 1 not present in the array, so we take any one of them, say 0010, and construct the third row by adding 0010 to the vectors in the top row.
- There is still one vector, 0100, of weight 1, not listed; we are forced to make it the coset leader of the last row and construct the last row as above. Due to the choices made, there is more than one correct answer but the weights of the chosen coset leaders must be $0, 1, 1, 1$.

To give the required example, choose any codevector, for example $\underline{c} = 0000$. We need to find a received vector $\underline{y}$ which differs from $\underline{c}$ in one bit but is not in the column of $\underline{c}$. Such a vector is $\underline{y} = 1000$. Thus, 0000 is transmitted, 1000 is received, $\mathrm{Decode}(\underline{y}) = 1001 \neq 0000$ so that the error is not corrected. (*Examples may vary.*) $\boxed{2\text{m}}$

(d) $\boxed{\text{ILO3, medium difficulty}}$ Assume that $\underline{y}, \underline{z} \in \mathbb{F}_2^4$, $\underline{y} \neq \underline{z}$, and that H is a parity check matrix for the code D such that $\underline{y}H^T = \underline{z}H^T$. Can $\underline{y}$ and $\underline{z}$ occur in the same column of the standard array that you constructed in part (c)? Justify your answer briefly.

Answer. [*unseen but an easy rider — 3*]

No, such vectors cannot occur in the same column of any standard array. Indeed, let a parity check matrix H be chosen. Then $S(\underline{x}) = \underline{x}H^T$ defines a syndrome map $S\colon \mathbb{F}_2^4 \to \mathbb{F}_2^2$. By a result from the course, $S(\underline{y}) = S(\underline{z})$ if and only if $\underline{y}$ and $\underline{z}$ belong to the same coset of D. But this would mean that $\underline{y}$ and $\underline{z}$ are in the same row of the standard array, hence not in the same column. $\boxed{3\text{m}}$

[20 marks]

A3.

(a) $\boxed{\text{ILO1, ILO2, basic}}$ Define what is meant by a *line* in the vector space $\mathbb{F}_q^r$. Explain how to construct a check matrix H for a Hamming $\mathrm{Ham}(r, q)$-code. State without proof the number of rows and the number of columns of H.

Answer. [*bookwork — 5*]

A line in $\mathbb{F}_q^r$ is a one-dimensional subspace of $\mathbb{F}_q^r$. $\boxed{\mathbf{1}\text{m}}$

A check matrix H has r rows and $(q^r - 1)/(q - 1)$ columns $\boxed{\mathbf{2}\text{m}}$ and its columns are (non-zero) representative vectors of lines in $\mathbb{F}_q^r$, one representative from each line. $\boxed{\mathbf{2}\text{m}}$

(b) $\boxed{\text{ILO5, basic}}$ Write down a check matrix of a ternary $\mathrm{Ham}(2, 3)$-code.

Answer. [*seen — 2*]

For example, $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}$. $\boxed{\mathbf{2}\text{m}}$

Comment. The columns of the matrix must represent the four lines in $\mathbb{F}_3^4$. That is, there must be four columns, and no two columns must be proportional.

(c) $\boxed{\text{ILO5, difficult}}$ Write down a check matrix of a $\mathrm{Ham}(2, 5)$-code $C \subseteq \mathbb{F}_5^6$ such that C contains the vector 111111.

Answer. [*Hamming code check matrices were done in class but without the zero sum constraint — 3*]

An example of a valid answer is $\begin{bmatrix} 1 & 3 & 1 & 1 & 4 & 0 \\ 1 & 1 & 3 & 4 & 0 & 1 \end{bmatrix}$. $\boxed{\mathbf{3}\text{m}}$

Comment. How to obtain such a matrix? One can pick any check matrix for $\mathrm{Ham}(2, 5)$ where the 6 columns represent all lines in $\mathbb{F}_5^2$, such as $H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 & 1 \end{bmatrix}$ (seen in the course). One checks whether the code defined by H contains 111111, meaning $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} H^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, or, the same, the sum of columns of H is zero. If this is not the case, a matrix of a linearly equivalent code should be obtained by scaling some columns of H. In the example of H above, it quickly becomes clear that the sum of columns cannot be made zero by scaling just one column, but a suitable matrix can be obtained by scaling, say, two columns.

[10 marks]

SECTION B

Answer **TWO** of the three questions in this section (40 marks in total).
If more than TWO questions from this section are attempted, then credit will be given for the best TWO answers.

B4. In this question, F is a finite alphabet of cardinality q, and $C \subseteq F^n$ is a code of cardinality $M > 1$.

(a) $\boxed{\text{ILO1, basic}}$ Define what is meant by the *minimum distance*, $d(C)$, of the code C.

Answer. [*bookwork — 2*]

$d(C) = \min\{d(\underline{x}, \underline{y}) : \underline{x}, \underline{y} \in C,\ \underline{x} \neq \underline{y}\}$ where $d(\underline{x}, \underline{y})$ denotes the Hamming distance between the words $\underline{x}$ and $\underline{y}$. $\boxed{2\text{m}}$

(b) $\boxed{\text{ILO1, medium difficulty}}$ Prove the Singleton bound which says that $M \leqslant q^{n-d(C)+1}$.

Answer. [*bookwork — 5*]

Denote $d(C)$ by d. Consider the function $f\colon C \to F^{n-d+1}$ where $f(\underline{v})$ is the word obtained from $\underline{v}$ by deleting the last $d-1$ symbols.

The function f is injective: indeed, if $\underline{v}, \underline{w} \in C$, $\underline{v} \neq \underline{w}$, then by definition of the minimum distance, $\underline{v}$ and $\underline{w}$ differ in at least d positions. Since f deletes only $d-1$ symbol, the words $f(\underline{v})$ and $f(\underline{w})$ still differ in at least one position. So $f(\underline{v}) \neq f(\underline{w})$.

Now, by the Pigeonhole Principle, injective functions $f\colon C \to F^{n-d+1}$ exist only if $\#C \leq \#F^{n-d+1}$. We conclude that $M \leq q^{n-d+1}$. $\boxed{5\text{m}}$

(c) $\boxed{\text{ILO1, basic}}$ Let r be an integer between 0 and n. Define the *Hamming sphere* $S_r(\underline{u})$ of radius r centred at $\underline{u} \in F^n$. State and prove the formula for the cardinality of $S_r(\underline{u})$ in terms of r, n and q.

Answer. [*bookwork — 5*]

The Hamming sphere $S_r(\underline{u})$ is the set $\{\underline{y} \in F^n : d(\underline{y}, \underline{u}) \leq r\}$. $\boxed{1\text{m}}$

One has $\#S_r(\underline{u}) = \sum_{i=0}^{r} \binom{n}{i}(q-1)^i$. $\boxed{2\text{m}}$

Observe that to choose a word at distance i from the word $\underline{u}$, one needs to select i positions where the word will differ from $\underline{u}$ — this can be done in $\binom{n}{i}$ ways — then in each of the chosen i positions, select a symbol different from the symbol in $\underline{u}$, for which there are $q-1$ ways. This shows that $\#\{\underline{y} : d(\underline{y}, \underline{u}) = i\} = \binom{n}{i}(q-1)^i$. Taking the sum from $i = 0$ to r gives the cardinality of $S_r(\underline{u})$ as stated. $\boxed{2\text{m}}$

(d) $\boxed{\text{ILO1, basic}}$ State without proof the Hamming bound on M in terms of n, q and $d(C)$. Define what is meant by saying that C is a *perfect code*.

Answer. [*bookwork —3*]

The Hamming bound is the inequality $M \leq q^n / \#S_t(\underline{u})$ for $t = [(d(C)-1)/2]$ and arbitrary $\underline{u}$ (*accept equivalent statements*). $\boxed{2\text{m}}$

The code C is perfect if the Hamming bound is attained, i.e., holds with equality. $\boxed{1\text{m}}$

(e) $\boxed{\text{ILO6 calculus, difficult}}$ Show that if $n = q = 60$ and C is perfect, then $d(C) \neq 7$. Hint: you may use the Singleton bound.

Answer. [*unseen — 5*]

Assume for contradiction that $d(C) = 7$. Then $t = [\frac{7-1}{2}] = 3$, and the denominator of the Hamming bound is

$$\#S_3(\underline{u}) = \sum_{i=0}^{3} \binom{n}{i}(q-1)^i = \binom{q}{0} + \binom{q}{1}(q-1) + \binom{q}{2}(q-1)^2 + \binom{q}{3}(q-1)^3. \qquad \boxed{2m}$$

Let us show that this denominator is less than q^6. Observe that $\binom{q}{3} = q(q-1)(q-2)/6 < q^3/6$ and that $(q-1)^3 < q^3$. Hence the last term of the denominator is less than $q^6/6$. We will also use the inequalities $\binom{q}{2}(q-1)^2 < q^4$ and $\binom{q}{1}(q-1) < q^2$. It follows that the whole denominator of the Hamming bound is less than $1 + q^2 + q^4 + (q^6/6)$ which is clearly less than q^6, given that q is 60.

Since C is perfect, M equals the Hamming bound, so $M > q^n/q^6 = q^{n-6}$. Yet by the Singleton bound, $M \leq q^{n-d(C)+1} = q^{n-6}$. This contradiction shows that the assumption $d(C) = 7$ was false. $\boxed{3m}$

Remark. The Tietäväinen–van Lint theorem about classification of perfect codes, seen in the course, does not help in this question because q is not a prime power.

[20 marks]

B5. In this question, $C \subseteq \mathbb{F}_q^n$ is a linear code of dimension k.

(a) $\boxed{\text{ILO1, basic}}$ Define what is meant by the *inner product* $\underline{x} \cdot \underline{y}$ of two vectors $\underline{x}, \underline{y} \in \mathbb{F}_q^n$, and what is meant by the *dual code* $C^\perp$. State without proof the length and the dimension of $C^\perp$.

Answer. [*bookwork — 5*]

The inner product of $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$ is the scalar $\underline{x} \cdot \underline{y} = x_1y_1 + \cdots + x_ny_n$. (*Accept* $\underline{x}\underline{y}^T$.) $\boxed{1m}$

The dual code is $C^\perp = \{\underline{y} \in \mathbb{F}_q^n : \underline{y} \cdot \underline{c} = 0,\ \forall \underline{c} \in C\}$. $\boxed{2m}$

The code $C^\perp$ has length n and dimension $n - k$. $\boxed{2m}$

(b) $\boxed{\text{ILO3, medium difficulty}}$ Recall that C is *self-orthogonal* if $C \subseteq C^\perp$. Prove that if C has a generator matrix G such that GG^T is a zero matrix, then C is self-orthogonal.

Answer. [*seen — 4*]

We need to show that if $\underline{y} \in C$ then $\underline{y} \cdot \underline{c} = 0$ for all $\underline{c} \in C$. Since G is a generator matrix, every codevector of C can be written as $\mathrm{Encode}(\underline{u}) = \underline{u}G$ for some $\underline{u} \in \mathbb{F}_q^k$. $\boxed{2m}$

Hence we are proving that $(\underline{u}G) \cdot (\underline{v}G) = 0$ for all $\underline{u}, \underline{v} \in \mathbb{F}_q^k$, equivalently $(\underline{u}G)(\underline{v}G)^T = 0$ which is $\underline{u}GG^T\underline{v} = 0$. But the latter is true given that GG^T is zero. $\boxed{2m}$

(c) $\boxed{\text{ILO2, ILO6 number theory, medium difficulty}}$ A self-orthogonal code $D \subseteq \mathbb{F}_q^4$ has generator matrix $\begin{bmatrix} 1 & 2 & -1 & 2 \\ 2 & 1 & 2 & -1 \end{bmatrix}$. Find a prime number $q > 2$ for which this is possible. Then write down a check matrix for D, explaining how it is obtained.

Answer. [*unseen — 4*]

Let us find a prime $q > 2$ for which $GG^T = 0$; then by part (b) the code is self-orthogonal. We have $GG^T = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$ which is a zero matrix when $q = 5$. $\boxed{\textbf{2m}}$

The quickest way to obtain a check matrix for D is to note that $D \subseteq D^\perp$ and $\dim D = \dim D^\perp = 2$ so $D = D^\perp$ and the matrix given in the question is therefore also a check matrix for D. $\boxed{\textbf{2m}}$

However, an alternative (and less elegant in this case) way to obtain a check matrix is to bring the generator matrix to standard form then apply a result from the course. Since $q = 5$, we have

$$\begin{bmatrix} 1 & 2 & -1 & 2 \\ 2 & 1 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & 2 & 4 \end{bmatrix} \xrightarrow{r_2 += 3r_1} \begin{bmatrix} 1 & 2 & 4 & 2 \\ 0 & 2 & 4 & 0 \end{bmatrix} \xrightarrow{r_1 -= r_2, r_2 \times = 2^{-1}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \end{bmatrix} = [I_2|A]$$

so by a result from the course, a possible check matrix for D is $[-A^T|I_2] = \begin{bmatrix} 0 & 3 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{bmatrix}$.

(d) $\boxed{\text{ILO6 linear algebra and combinatorics, difficult}}$ Assume that $q = 2$ and $C \subseteq C^\perp$. Let $W_C(x, y) = \sum_{i=0}^n A_i x^{n-i} y^i$ be the weight enumerator of C. Prove that $A_2 \leqslant k$ and that $A_2(A_2 - 1)/2 \leqslant A_4$.

Answer. [*unseen — 7*]

Recall that, by definition of the weight enumerator, A_2 is the number of codevectors of weight 2 in C. Let $e_{ij} \in \mathbb{F}_2^n$ be the vector of weight 2 where the ith and the jth bit are 1. Note that if C contains e_{ij} and e_{st}, then $e_{ij} \cdot e_{st} = 0$ which means that either $\{i, j\} = \{s, t\}$ or $\{i, j\} \cap \{s, t\} = \varnothing$. That is, vectors of weight 2 in C correspond to disjoint pairs of indices; hence the set of vectors of weight 2 in C is linearly independent. It immediately follows that $A_2 \leq \dim C = k$. $\boxed{\textbf{4m}}$

Furthermore, the above implies that any pair of codevectors of weight 2 sum to a codevector of weight 4, and all such $\binom{A_2}{2}$ sums are distinct, thereby creating $\binom{A_2}{2}$ codevectors of weight 4. Therefore, A_4, which is the total number of codevectors of weight 4, is at least $\binom{A_2}{2} = A_2(A_2 - 1)/2$, as claimed. $\boxed{\textbf{3m}}$

[20 marks]

B6. In this question, we assume that vectors in $\mathbb{F}_q^n$ are identified with polynomials, in $\mathbb{F}_q[x]$, of degree less than n in the usual way.

(a) $\boxed{\text{ILO1, medium difficulty}}$ Let $C \subseteq \mathbb{F}_q^n$ be a cyclic code. Explain what is meant by a *generator polynomial* of C. Show that there exists at most one generator polynomial of C.

Answer. [*bookwork — 5*]

A generator polynomial $g(x)$ is an element of C $\boxed{\textbf{1m}}$ such that $g(x)$ is monic $\boxed{\textbf{1m}}$ and every polynomial in C is a multiple, in $\mathbb{F}_q[x]$, of $g(x)$. $\boxed{\textbf{1m}}$ (*Accept equivalent criteria such as $g(x)$ is monic of least degree in C.*)

Assume that $g_1(x)$ and $g_2(x)$ are generator polynomials of C. Then by definition $g_2(x) \in C$ so, again by definition, $g_1(x)$ divides $g_2(x)$. In the same way, $g_2(x)$ divides $g_1(x)$. Since both polynomials are monic, this implies that $g_1(x) = g_2(x)$, proving uniqueness. $\boxed{\textbf{2m}}$

(b) $\boxed{\text{ILO2, medium difficulty}}$ Let $Z \subseteq \mathbb{F}_7^8$ be the cyclic code with generator polynomial $x - 1$. Write down a generator matrix for Z. Prove that $Z = \{(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8) \in \mathbb{F}_7^8 : \sum_{i=1}^8 v_i = 0 \text{ in } \mathbb{F}_7\}$.

Answer. [*routine method + proof which was seen for other n and q — 5*]

The generator polynomial $g(x) = -1 + x$ corresponds to the vector $\begin{bmatrix}-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$ in $\mathbb{F}_7^8$, so a standard construction gives the following 7×8 generator matrix for Z:

$$G = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}. \qquad \boxed{\text{2m}}$$

Denote by Y the code $\{(v_1, \ldots, v_8) \in \mathbb{F}_7^8 : \sum_{i=1}^8 v_i = 0\}$, given in the question (known as the *zero sum code* of length 8). To prove that $Z = Y$, one may observe that each row of above generator matrix G has zero sum hence belongs to Y. Since Z is spanned by the rows of G, we have $Z \subseteq Y$. However, $\dim Z = 7$ as G has 7 rows, and $\dim Y \le 7$ as $Y \ne \mathbb{F}_7^8$, so $\dim Z \ge \dim Y$, and as $Z \subseteq Y$, we have $Z = Y$. $\boxed{\text{3m}}$

You may use the factorisation $(x-1)(x+1)(x^2+1)(x^2-3x+1)(x^2+3x+1)$ of $x^8 - 1$ into irreducible monic polynomials in $\mathbb{F}_7[x]$ to answer the following questions.

(c) $\boxed{\text{ILO5, medium difficulty}}$ Determine: (i) the total number of cyclic codes $C \subseteq \mathbb{F}_7^8$; (ii) the number of cyclic codes $C \subseteq \mathbb{F}_7^8$ such that $C \subseteq Z$ where Z is the code from part (b). Justify your answers briefly.

Answer. [*(i) standard, (ii) easy rider — 4*]

(i) The total number of cyclic codes in $\mathbb{F}_7^8$ is equal to the number of possible generator polynomials, which are monic polynomials dividing $x^8 - 1$. Such a polynomial is a product of a subset of the set of irreducible monic factors of $x^8 - 1$ given above. Since there are 5 factors, the number of such subsets is $2^5 = 32$. Answer: 32. $\boxed{\text{2m}}$

(ii) Since Z consists of all polynomials of degree < 8 divisible by $x - 1$, a cyclic code C is contained in Z if and only if all polynomials in C are divisible by $x - 1$. Equivalently, the generator polynomial $g(x)$ of C is divisible by $x - 1$. Hence $g(x)$ must be a product of $x - 1$ and a subset of the set of 4 remaining irreducible monic factors of $x^8 - 1$, which can be chosen in $2^4 = 16$ ways. Answer: 16. $\boxed{\text{2m}}$

(d) $\boxed{\text{ILO5, ILO2, difficult}}$ Prove that none of the cyclic codes in $\mathbb{F}_7^8$ is a Hamming code. Any facts from the course can be freely used.

Answer. [*unseen — 6*]

Assume for contradiction that a cyclic code $C \subseteq \mathbb{F}_7^8$ is Hamming. Then C must be $\mathrm{Ham}(2,7)$, because the length of $\mathrm{Ham}(r,7)$ is $1 + 7 + \cdots + 7^{r-1}$ which is greater than 8 unless $r = 2$. $\boxed{\text{1m}}$

Observe further that the dimension of $\mathrm{Ham}(2,7)$ is $8 - 2 = 6$. Hence the generator polynomial $g(x)$ of C must be of degree 2. $\boxed{\text{1m}}$

The monic divisors of $x^8 - 1$ of degree 2 are $(x-1)(x+1) = x^2 - 1$, $x^2 + 1$, $x^2 - 3x + 1$ and $x^2 + 3x + 1$. The first two polynomials correspond to vectors of weight 2 hence cannot be in C, as the weight of a Hamming code is 3. $\boxed{\text{1m}}$

This leaves us with two possibilities for $g(x)$: $x^2 - 3x + 1$ and $x^2 + 3x + 1$. Let us rule both of them out. Assume $g(x) = x^2 - 3x + 1$. Then the check polynomial of C is $h(x) = (x-1)(x+1)(x^2+1)(x^2+3x+1) = (x^4 - 1)(x^2 + 3x + 1) = x^6 + 3x^5 + x^4 - x^2 - 3x - 1$. The vector formed by the coefficients of $h(x)$ written backwards, $\begin{bmatrix}1 & 3 & 1 & 0 & -1 & -3 & -1 & 0\end{bmatrix}$, has weight 6 (accounting for the fact that x^3 is not present) and lies in $C^\perp$. Yet $C^\perp$ is a simplex code (a dual of the Hamming code) where, by a result from the course, every non-zero codevector must have weight $q^{r-1} = 7$ and not 6. This contradiction proves that C is not Hamming. (*Alternatively*, a 2×8 check matrix can be written down using $h(x)$, and two proportional columns can be pointed out, e.g., $[0\,1]^T$ and $[0\,-1]^T$.)

The case $g(x) = x^2 + 3x + 1$ is treated in the same way. $\boxed{\mathbf{3}\text{m}}$

[20 marks]

END OF EXAMINATION PAPER

Question Q1(i)

Learning Outcome

Classify Möbius transformations in terms of their actions on the hyperbolic plane: (i), (ii) low level; (iii), (iv) medium level; (v) high level.

(i) is bookwork. The definitions in (ii) are bookwork, the examples are similar to exercises. (iii) is from the exercise sheets. (iv) was sketched in the lectures with details left as an exercise. (v) is unseen.

Solution

(i) A Möbius transformation is a transformation of the form $\gamma(z) = (az+b)/(cz+d)$, $a, b, c, d \in \mathbb{R}$, $ad-bc > 0$. [2 marks]

(ii) Let $\gamma(z) = (az+b)/(cz+d)$. By dividing numerator and denominator by $\sqrt{ad-bc}$ there is no loss in generality in assuming that $ad-bc = 1$. When $ad-bc = 1$ we define $\tau(\gamma) = (a+d)^2$. [2 marks]

γ is hyperbolic precisely when $\tau(\gamma) > 4$.

γ is parabolic precisely when $\tau(\gamma) = 4$.

γ is elliptic precisely when $\tau(\gamma) \in [0, 4)$. [3 marks]

γ_1 is not normalised. In normalised form we have

$$\gamma_1(z) = \frac{\frac{7}{3}z - \frac{8}{3}}{\frac{2}{3}z - \frac{1}{3}}.$$

Hence $\tau(\gamma_1) = (7/3 - 1/3)^2 = 4$ so that γ_1 is parabolic.

γ_2 is not normalised. In normalised form we have

$$\gamma_2(z) = \frac{\frac{2}{3}z - \frac{1}{3}}{\frac{-7}{3}z + \frac{8}{3}}.$$

Hence $\tau(\gamma_1) = (2/3 + 8/3)^2 = \frac{100}{9} > 4$ so that γ_1 is hyperbolic. [4 marks]

(iii) Let $\gamma_1 = (a_1z+b_1)/(c_1z+d_1)$ and $\gamma_2 = (a_2z+b_2)/(c_2z+d_2)$. Then their composition is

$$\begin{aligned}\gamma_2\gamma_1(z) &= \frac{a_2\left(\frac{a_1z+b_1}{c_1z+d_1}\right)+b_2}{c_2\left(\frac{a_1z+b_1}{c_1z+d_1}\right)+d_2} \\ &= \frac{(a_2a_1+b_2c_1)z+(a_2b_1+b_2d_1)}{(c_2a_1+d_2c_1)z+(c_2b_1+d_2d_1)},\end{aligned}$$

which is a Möbius transformation of $\mathbb{H}$ as

$$\begin{aligned}&(a_2a_1+b_2c_1)(c_2b_1+d_2d_1)-(a_2b_1+b_2d_1)(c_2a_1+d_2c_1) \\ &= (a_1d_1-b_1c_1)(a_2d_2-b_2c_2) > 0.\end{aligned}$$

[Full credit will also be given if an explanation using the connection between composition of Möbius transformations and multiplication of matrices is given.] [6 marks]

(iv) Suppose that z satisfies $\alpha z\bar{z} + \beta z + \beta\bar{z} + \gamma = 0$. Let $w = \gamma(z)$ where $\gamma(z) = (az+b)/(cz+d)$. Then $z = (dw-b)/(-cw+a)$. Hence w satisfies

$$\alpha\frac{dw-b}{-cw+a}\frac{d\bar{w}-b}{-c\bar{w}+a} + \beta\frac{dw-b}{-cw+a} + \beta\frac{d\bar{w}-b}{-c\bar{w}+a} + \gamma = 0$$

i.e.

$$\alpha(dw-b)(d\bar{w}-b)+\beta(dw-b)(-c\bar{w}+a)+\beta(d\bar{w}-b)(-cw+a)+\gamma(-cw+a)(-c\bar{w}+a) = 0$$

Hence

$$\alpha' w\bar{w} + \beta' w + \beta'\bar{w} + \gamma' = 0$$

with

$$\begin{aligned}\alpha' &= \alpha d^2 - 2\beta dc + \gamma c^2 \\ \beta' &= -\alpha bd + \beta ad + \beta bc - \gamma ac \\ \gamma' &= \alpha b^2 - 2\beta ab + \gamma a^2\end{aligned}$$

which has the same form as above. [8 marks]

(v) By the fact, γ maps circles to straight lines or to circles. Moreover γ maps $\partial\mathbb{H}$ to $\partial\mathbb{H}$. Note that A intersects $\partial\mathbb{H}$ at exactly one point. Hence $\gamma(A)$ must be either a straight line or a circle that intersects $\partial\mathbb{H}$ at exactly one point. In the first case, $\gamma(A)$ must be a horizontal straight line (which intersects $\partial\mathbb{H}$ at ∞ only; any other straight line will meet $\partial\mathbb{H}$ at two points). In the second case, $\gamma(A)$ must be a circle that touches the real axis at a single point. [5 marks]

Question Q2

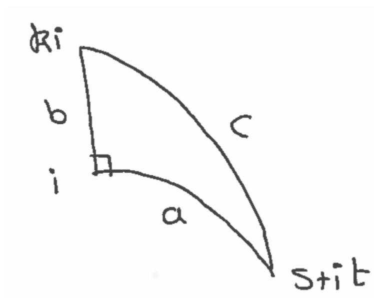
Learning Outcome

Prove results (Gauss-Bonnet Theorem, angle formulæ for triangles, etc as listed in the syllabus) in hyperbolic trigonometry and use them to calculate angles, side lengths, hyperbolic areas, etc, of hyperbolic triangles and polygons. (i), (ii) medium level. (iii) low level. (iv) high level.

(i) is bookwork. (ii) is from the exercise sheets. (iii) is bookwork. The first part of (iv) is a particular case of a result discussed in lectures; the second part is unseen.

Solution

(i) By applying a Möbius transformation of $\mathbb{H}$, we may assume that the vertex with internal angle $\pi/2$ is at i and that the side of length b lies along the imaginary axis; here we are using the fact that Möbius transformations are conformal. It follows that the side of length a lies along the geodesic given by the semi-circle centred at the origin with radius 1. Therefore, the other vertices of Δ can be taken to be at ki for some $k > 0$ and at $s + it$, where $s + it$ lies on the circle centred at the origin and of radius 1.



Using the formula for $\cosh d_{\mathbb{H}}(z, w)$ we have

$$\cosh a = 1 + \frac{|s + i(t-1)|^2}{2t} = 1 + \frac{s^2 + (t-1)^2}{2t} = \frac{1}{t}, \qquad (1)$$

$$\cosh b = 1 + \frac{(k-1)^2}{2k} = \frac{1+k^2}{2k}, \qquad (2)$$

$$\cosh c = 1 + \frac{|s + i(t-k)|^2}{2tk} = 1 + \frac{s^2 + (t-k)^2}{2tk} = \frac{1+k^2}{2tk}, \qquad (3)$$

(using the fact that $s^2 + t^2 = 1$). Hence $\cosh c = \cosh a \cosh b$. [8 marks]

(ii) Construct a geodesic from vertex B to the geodesic segment $[A, C]$ in such a way that these geodesics meet at right-angles. This splits Δ into two right-angled triangles, BDA and BDC. Let the length of the geodesic segment $[B, D]$ be d, and suppose that BDA has internal angles $\beta_1, \pi/2, \alpha$ and side lengths d, b_1, c. Label BDC similarly.

We know that

$$\sin\beta_1 = \frac{\sinh b_1}{\sinh c},\ \cos\beta_1 = \frac{\tanh d}{\tanh c},\ \sin\beta_2 = \frac{\sinh b_2}{\sinh a},\ \cos\beta_2 = \frac{\tanh d}{\tanh a}.$$

By the hyperbolic version of Pythagoras' Theorem we know that

$$\cosh c = \cosh b_1 \cosh d,\ \cosh a = \cosh b_2 \cosh d.$$

Hence

$$\begin{aligned}
\sin\beta &= \sin(\beta_1+\beta_2) \\
&= \sin\beta_1\cos\beta_2+\sin\beta_2\cos\beta_1 \\
&= \frac{\sinh b_1}{\sinh c}\frac{\sinh d}{\cosh d}\frac{\cosh a}{\sinh a}+\frac{\sinh b_2}{\sinh a}\frac{\sinh d}{\cosh d}\frac{\cosh c}{\sinh c} \\
&= \frac{\sinh b_1 \sinh d}{\sinh c \sinh a}\cosh b_2+\frac{\sinh b_2 \sinh d}{\sinh a \sinh c}\cosh b_1 \\
&= \frac{\sinh d}{\sinh a \sinh c}(\sinh b_1\cosh b_2+\sinh b_2\cosh b_1) \\
&= \frac{\sinh d}{\sinh a \sinh c}\sinh(b_1+b_2) \\
&= \frac{\sinh b \sinh d}{\sinh a \sinh c}.
\end{aligned}$$

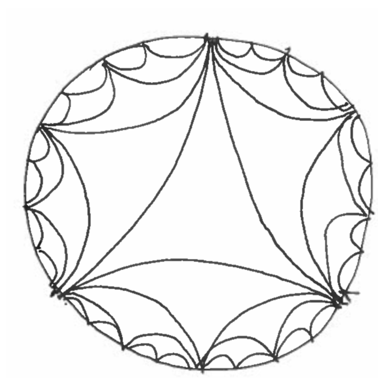
Hence $\sin\alpha = \sinh d/\sinh c$ and $\sin\gamma = \sinh d/\sin a$. Substituting these into the above equality proves the result. [12 marks]

(iii) Let Δ be a hyperbolic triangle with internal angles α, β and γ. Then $\text{Area}_{\mathbb{H}} = \pi-(\alpha+\beta+\gamma)$. [2 marks]

(iv) Suppose the hyperbolic triangle Δ has internal angle α. As k polygons meet at each vertex, we must have that $k\alpha = 2\pi$, i.e. $\alpha = 2\pi/k$. By the Gauss-Bonnet Theorem

$$0 \le \text{Area}_{\mathbb{H}}(P) = \pi - 3\times\frac{2\pi}{k}.$$

Hence $1-6/k>0$, i.e. $k>6$. Hence $k\geq 7$. [4 marks]



[4 marks]

Question Q3

Learning Outcome	**Solution**

Calculate a fundamental domain and a set of side-pairing transformations for a given Fuchsian group. (a)(i)–(iii) at low level. (b)(i), (ii), (iv) at medium level.

Compare different models (the upper half-plane model and the Poincaré disc model) of hyperbolic geometry. (b)(iii) at medium level.

(a)(i), (a)(ii) are bookwork. (a)(iii) is from the exercise sheets. (b)(i) is unseen. (b)(ii), (b)(iii), (b)(iv) are similar to exercise

(a) (i) $F\subset\mathbb{H}$ is a fundamental domain for Γ if (i) $\bigcup_{\gamma\in\Gamma}\gamma(\overline{F}) = \mathbb{H}$, (ii) $\gamma_1(F)\cap\gamma_2(F)=\emptyset$ if $\gamma_1,\gamma_2\in\Gamma$, $\gamma_1\neq\gamma_2$. [2 marks]

(ii) Choose a point $p\in\mathbb{H}$ such that $\gamma(p)\neq p$ for all $\gamma\in\Gamma\setminus\{\text{id}\}$.
For each $\gamma\in\Gamma\setminus\{\text{id}\}$, construct the arc of geodesic $[p,\gamma(p)]$ from p to $\gamma(p)$.
Let $L_p(\gamma)$ denote the perpendicular bisector of $[p,\gamma(p)]$. Then $L_p(\gamma)$ divides $\mathbb{H}$ into two half-planes. One of these half-planes contains p; call this half-plane $H_p(\gamma)$.
Then $D(p) := \bigcap_{\gamma\in\Gamma\setminus\{\text{id}\}} H_p(\gamma)$ is a Dirichlet region for Γ. [6 marks]

(iii) Write $z_1 = x_1+iy_1, z_2 = x_2+iy_2$. Then $z=x+iy$ is on the perpendicular bisector of $[z_1,z_2]$ if and only if $d_{\mathbb{H}}(z,z_1) = d_{\mathbb{H}}(z,z_2)$ if and only if $\cosh d_{\mathbb{H}}(z,z_1) = \cosh d_{\mathbb{H}}(z,z_2)$ if and only if

$$1+\frac{|z-z_1|^2}{2\,\text{Im}\,z\,\text{Im}\,z_1} = 1+\frac{|z-z_2|^2}{2\,\text{Im}\,z\,\text{Im}\,z_2},$$

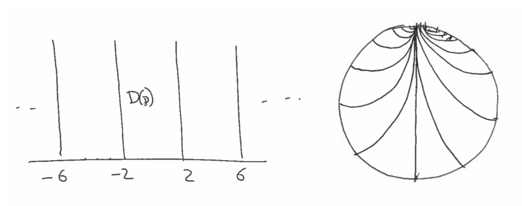
sheets.

i.e.
$$\frac{|z-z_1|^2}{2yy_1} = \frac{|z-z_2|^2}{2yy_2}.$$

Simplifying this gives the claimed expression. [4 marks]

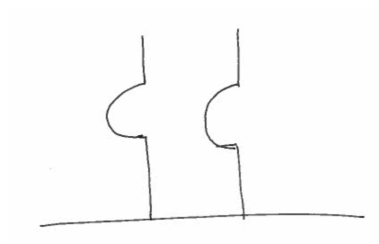
(b) (i) Let $z = x+iy$ be on the perpendicular bisector of $[i, i+b]$. Then
$$|(x+iy)-i|^2 = |(x+iy)-(i+b)|^2$$
i.e.
$$\begin{aligned} x^2+(y-1)^2 &= (x-b)^2+(y-1)^2 \\ x^2 &= x^2-2xb+b^2 \\ 2xb &= b^2 \\ x &= b/2, \end{aligned}$$
so the perpendicular bisector is the vertical straight line with real part $b/2$. [4 marks]

(ii) Let $\gamma_n(z) = z+4n$. Let $p = i$. By (ii) we have that $L_p(\gamma_n)$ is the vertical straight line with real part $2n$. Hence
$$H_p(\gamma_n) = \begin{cases} \{z \in \mathbb{H} \mid \mathrm{Re}(z) < 2n\} \text{ if } n \geq 0 \\ \{z \in \mathbb{H} \mid \mathrm{Re}(z) > 2n\} \text{ if } n \leq 0. \end{cases}$$
Hence
$$D(p) = \bigcap_{n\neq 0} H_p(\gamma_n) = \{z \in \mathbb{H} \mid -2 < \mathrm{Re}(z) < 2\}.$$
[6 marks]

(iii) 

[4 marks]

(iv) $D(p)$ for an arbitrary choice of $p \in \mathbb{H}$ will be a vertical strip of width 4.
An example of a fundamental domain for Γ that is not of the form $D(p)$ for some $p \in \mathbb{H}$ is illustrated below.



[Any similar repeating shape that isn't a vertical strip is acceptable.] [4 marks]

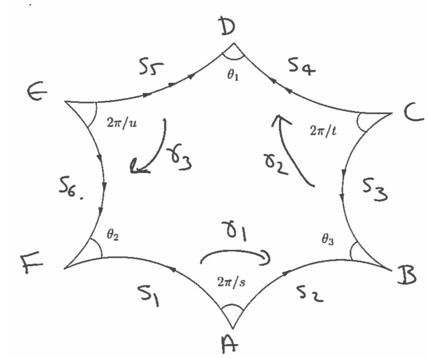
Question Q4

Learning Outcome	**Solution**
Use Poincaré's Theorem to construct examples of Fuchsian groups and calculate presentations in terms of generators and relations. (i) at low level. (ii) at medium level. (iii) at high level.	(i) The angle sum is $\mathrm{sum}(\mathcal{E}) := \sum_{j=1}^{n} \angle v_j$. $\mathcal{E}$ satisfies the elliptic cycle condition if there exists an integer $n \geq 1$ such that $\mathrm{sum}(\mathcal{E}) = 2\pi/n$. $\mathcal{E}$ is an accidental cycle if $\mathrm{sum}(\mathcal{E}) = 2\pi$. [4 marks]

Relate the signature of a Fuchsian group to the algebraic and geometric properties of the Fuchsian group and to the geometry of the corresponding hyperbolic surface.. (iv) definition at low level, remainder at high level.

(i) is bookwork; (ii), (iii) are similar to exercise sheets. The definitions in (iv) are bookwork, the remainder is unseen.

(ii) Label the diagram as below.



The elliptic cycles are:

(1)

$$\begin{pmatrix} A \\ s_1 \end{pmatrix} \xrightarrow{\gamma_1} \begin{pmatrix} A \\ s_2 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} A \\ s_1 \end{pmatrix}$$

elliptic cycle: $\mathcal{E}_1 = A$.
elliptic cycle transformation: γ_1.
angle sum: $\text{sum}(\mathcal{E}_1) = 2\pi/n_1$.

(2)

$$\begin{pmatrix} C \\ s_3 \end{pmatrix} \xrightarrow{\gamma_2} \begin{pmatrix} C \\ s_4 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} C \\ s_3 \end{pmatrix}$$

elliptic cycle: $\mathcal{E}_2 = C$.
elliptic cycle transformation: γ_2.
angle sum: $\text{sum}(\mathcal{E}_2) = 2\pi/n_2$.

(3)

$$\begin{pmatrix} E \\ s_5 \end{pmatrix} \xrightarrow{\gamma_3} \begin{pmatrix} E \\ s_6 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} E \\ s_5 \end{pmatrix}$$

elliptic cycle: $\mathcal{E}_3 = E$.
elliptic cycle transformation: γ_3.
angle sum: $\text{sum}(\mathcal{E}_3) = 2\pi/n_3$.

(4)

$$\begin{array}{cccc} \begin{pmatrix} B \\ s_3 \end{pmatrix} & \xrightarrow{\gamma_2} & \begin{pmatrix} D \\ s_4 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} D \\ s_5 \end{pmatrix} \\ & \xrightarrow{\gamma_3} & \begin{pmatrix} F \\ s_6 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} F \\ s_1 \end{pmatrix} \\ & \xrightarrow{\gamma_1} & \begin{pmatrix} B \\ s_2 \end{pmatrix} \xrightarrow{*} \begin{pmatrix} B \\ s_3 \end{pmatrix} \end{array}$$

elliptic cycle: $\mathcal{E}_4 = B \to D \to F$.
elliptic cycle transformation: $\gamma_1\gamma_3\gamma_2$.
angle sum: $\text{sum}(\mathcal{E}_4) = \theta_1 + \theta_2 + \theta_3$.

(iii) The elliptic cycle condition holds for $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ when s, t, u are integers, say $s = n_1, t = n_2, u = n_3$. Note that the question implies that $n_1, n_2, n_3 \geq 2$.

If $\theta_1 + \theta_2 + \theta_3 = 2\pi/n$ for some $n \geq 1$ then the elliptic cycle condition holds for $\mathcal{E}_4$ with order n.

Hence $\gamma_1, \gamma_2, \gamma_3$ generate a Fuchsian group Γ and

$$\Gamma = \langle a, b, c \mid a^{n_1} = b^{n_2} = c^{n_3} = (acb)^n = e \rangle.$$

[6 marks]

(iv) Let Γ be a cocompact Fuchsian group. Let $\mathcal{E}_1, \ldots, \mathcal{E}_r$ be the non-accidental elliptic cycles and let $\mathcal{E}_{r+1}, \ldots, \mathcal{E}_s$ be the accidental cycles. Let g be the genus of $\mathbb{H}/\Gamma$. Then $\mathrm{sig}(\Gamma) = (g; m_1, \ldots, m_r)$. (We write $(-; m_1, \ldots, m_r)$ if $g = 0$.)

In the example above, the hexagon forms a triangulation of $\mathbb{H}/\Gamma$ with $V = 4$ vertices (=number of elliptic cycles), $E = 3$ sides, and $F = 1$ faces. Hence by Euler's formula

$$2 - 2g = \chi(\mathbb{H}/\Gamma) = V - E + F = 4 - 3 + 1 = 2.$$

Hence $g = 0$.

Hence

$$\mathrm{sig}(\Gamma) = \begin{cases} (-; n_1, n_2, n_3, n) \text{ if } n \neq 1 \\ (-; n_1, n_2, n_3) \text{ if } n = 1. \end{cases}$$

In both cases, $\mathbb{H}/\Gamma$ is a topological sphere with 3 (in the first case) or 4 (in the second case) marked points. [8 marks]

MATH32062 solutions 2018–2019

* indicates material suitable for weaker students.

1. (a)(i) [1 mark, *, ILO 1 at low level]
$\mathcal{V}(J) = \{(a_1, a_2, \dots, a_n) \in K^n \mid f(a_1, a_2, \dots, a_n) = 0, \forall f \in J\}$
(ii) [2 marks, *, ILO 1 at low level]
$\mathcal{I}(V) = \{f \in K[x_1, x_2, \dots, x_n] \mid f(a_1, a_2, \dots, a_n) = 0, \forall (a_1, a_2, \dots, a_n) \in V\}$
$K[V] = K[x_1, x_2, \dots, x_n]/\mathcal{I}(V)$
(iii) [3 marks, *, ILO 4 at medium level]
If $P = (x_0, y_0, z_0) \in \mathcal{V}(J)$, then $(x_0 - y_0 + z_0)^2 = 0$, which implies $x_0 - y_0 + z_0 = 0$, therefore $x - y + z \in \mathcal{I}(\mathcal{V}(J))$. (It is also acceptable to write that $x - y + z \in \sqrt{J}$ as $(x - y + z)^2 \in \sqrt{J}$, and $\sqrt{J} \subseteq \mathcal{I}(\mathcal{V}(J))$, which was proved as part of the Nullstellensatz, holds without any condition on the field.)
Both generators of J are homogeneous of degree 2, therefore J does not contain any non-0 polynomial of degree less than 2, in particular, $x - y + z \notin J$. This means that $x - y + z \in \mathcal{I}(\mathcal{V}(J)) \setminus J$, so $\mathcal{I}(\mathcal{V}(J)) \neq J$.

(b)(i) [1 mark, *, ILO 1 at medium level]
W is reducible if and only if it can be written as $W = W_1 \cup W_2$, where W_1, W_2 are also affine algebraic varieties, $W_1 \neq W \neq W_2$. W is irreducible if and only if it is not reducible.
(ii) [3 marks, ILO 2 at high level, bookwork]
Assume that $\mathcal{I}(W)$ is not prime. Let f_1, f_2 be polynomials such that f_1, $f_2 \notin \mathcal{I}(W)$, but $f_1 f_2 \in \mathcal{I}(W)$. Let

$$W_1 = \{P \in W \mid f_1(P) = 0\} = W_1 \cap \mathcal{V}(\langle f_1 \rangle) = \mathcal{V}(\langle \mathcal{I}(W), f_1 \rangle)$$

and similarly

$$W_2 = \{P \in W \mid f_2(P) = 0\} = W_2 \cap \mathcal{V}(\langle f_2 \rangle) = \mathcal{V}(\langle \mathcal{I}(W), f_2 \rangle).$$

W_1, W_2 are affine algebraic varieties. Clearly $W_1 \subseteq W$, $W_2 \subseteq W$, but $W_1 \neq W$ since $f_1 \notin \mathcal{I}(W)$ and similarly $W_2 \neq W$ since $f_2 \notin \mathcal{I}(W)$. For any $P \in W$, $0 = (f_1 f_2)(P) = f_1(P)(f_2)(P)$, so $f_1(P) = 0$ or $f_2(P) = 0$. In the first case, $P \in W_1$, in the second $P \in W_2$, therefore $W = W_1 \cup W_2$ and W is reducible.
(iii) [5 marks, *, ILO 4 at high level, unseen, but similar to problems on problem sheet]
$xyz - xz = xz(y - 1)$, therefore by using the idea of the proof above just with 3 factors instead of 2, we can write $W = W_1 \cup W_2 \cup W_3$, where

$$W_1 = \mathcal{V}(\langle xyz - xz, x^2 + 2y^2 - z^2 - 2, x \rangle) = \mathcal{V}(\langle x^2 + 2y^2 - z^2 - 2, x \rangle),$$
$$W_2 = \mathcal{V}(\langle xyz - xz, x^2 + 2y^2 - z^2 - 2, z \rangle) = \mathcal{V}(\langle x^2 + 2y^2 - z^2 - 2, z \rangle),$$
$$W_3 = \mathcal{V}(\langle xyz - xz, x^2 + 2y^2 - z^2 - 2, y - 1 \rangle) = \mathcal{V}(\langle x^2 + 2y^2 - z^2 - 2, y - 1 \rangle)$$

$xyz - xz$ can be omitted in each case because it is divisible by the new generator. If we substitute $x = 0$ into $x^2 + 2y^2 - z^2 - 2 = 0$, we get $2y^2 - z^2 - 2 = 0$, which is the equation of a hyperbola, so W_1 is a hyperbola in the plane $x = 0$ and hyperbolas are known to be irreducible.
If we substitute $z = 0$ into $x^2 + 2y^2 - z^2 - 2 = 0$, we get $x^2 + 2y^2 - 2 = 0$, which is the equation of an ellipse, so W_1 is a ellipse in the plane $z = 0$ and ellipses are also known to be irreducible.
If we substitute $y = 1$ into $x^2 + 2y^2 - z^2 - 2 = 0$, we get $x^2 - z^2 = 0$, which can be factorised as $(x - z)(x + z) = 0$, so W_3 is the union of the lines $W_3' = \mathcal{V}(\langle y - 1, x - z\rangle)$ and $W_3'' = \mathcal{V}(\langle y - 1, x + z\rangle)$, which are irreducible as lines are irreducible.
Therefore $W = W_1 \cup W_2 \cup W_3' \cup W_3''$ and these four varieties are irreducible. The intersection of any two of them contains at most 2 points, while each of these varieties is an infinite set, so none of them contains any of the others, this also follows from their geometric descriptions, therefore W_1, W_2, W_3' and W_3'' are the irreducible components of W.

(c)(i) [2 marks, *, ILO 1 at medium level]
A *morphism* $\varphi : V \to W$ is a function $\varphi : V \to W$ of the form $\varphi(P) = (\varphi_1(P), \varphi_2(P), \ldots, \varphi_n(P))$ with $\varphi_i \in K[V]$ for each i, $1 \leq i \leq n$.
Let $f \in K[W]$. f is a function $W \to K$ and $\varphi^*(f) = f \circ \varphi$, which is a function $V \to K$.

(ii) [3 marks, ILO 4 at high level]
Let φ be a morphism $\mathbb{A}^1 \to C$. Then $\varphi = (\varphi_1, \varphi_2)$, where φ_1, $\varphi_2 \in \mathbb{R}[\mathbb{A}^1] = \mathbb{R}[t]$ and φ_1 and φ_2 satisfy $\varphi_1^2 + \varphi_2^2 - 1 = 0$. Let's assume φ_1 and φ_2 are not both constant. Let $d_1 = \deg \varphi_1$, $d_2 = \deg \varphi_2$, we can assume without loss of generality that $d_1 \geq d_2$. $d_1 \geq 1$ as φ_1 and φ_2 are not both constant, so the highest degree term in φ_1^2 is t^{2d_1} with a positive coefficient. If $d_1 = d_2$, then φ_2^2 also contains t^{2d_1} with a positive coefficient, if $d_1 > d_2$, then φ_2^2 has no t^{2d_1} term. In both cases $\varphi_1^2 + \varphi_2^2$ contains a t^{2d_1} term with a positive coefficient and $2d_1 \geq 2$, therefore $\varphi_1^2 + \varphi_2^2 - 1 \neq 0$, a contradiction.

2. (a)(i) [2 marks, *, ILO 3 at medium level]
ℓ is tangent to V at P if and only if $(\nabla f)_P.\mathbf{v} = 0$ for every $f \in \mathcal{I}(V)$.
The tangent space to V at P, denoted by T_PV is the union of the tangent lines to V at P and the point P.

(a)(ii) [5 marks, 1* mark for the criterion, ILO 3 at medium level, 4 marks for the proof, ILO 3 at high level, bookwork]
ℓ is tangent to V at P iff $J_P\mathbf{v} = 0$.
Assume that ℓ is tangent to V at P. Then $(\nabla f)_P.\mathbf{v} = 0$ for every $f \in \mathcal{I}(V)$, in particular, $(\nabla f_i)_P.\mathbf{v} = 0$ for every i, $1 \leq i \leq r$. $(\nabla f_i)_P.\mathbf{v}$ is exactly the ith component of $J_P\mathbf{v}$, therefore $J_P\mathbf{v} = 0$.

Assume now that $J_P\mathbf{v} = \mathbf{0}$. Let $f \in \mathcal{I}(V)$. Then $f = \sum_{i=1}^{r} g_i f_i$ for some $g_i \in K[x_1, x_2, \ldots, x_n]$, $1 \leq i \leq r$. Now

$$\nabla f = \sum_{i=1}^{r} \nabla(g_i f_i) = \sum_{i=1}^{r}((\nabla g_i) f_i + g_i \nabla f_i)$$

and so

$$(\nabla f)_P = \sum_{i=1}^{r}((\nabla g_i)_P f_i(P) + g_i(P)(\nabla f_i)_P) = \sum_{i=1}^{r} g_i(P)(\nabla f_i)_P.$$

In the last step we used that $f_i(P) = 0$ for every i, $1 \leq i \leq r$. Hence $(\nabla f)_P.\mathbf{v} = \sum_{i=1}^{r} g_i(P)(\nabla f_i)_P.\mathbf{v}$, but $(\nabla f_i)_P.\mathbf{v} = 0$, because it is the ith component of $J_P\mathbf{v} = \mathbf{0}$, therefore $(\nabla f)_P.\mathbf{v} = 0$.

(b)(i) [3 marks, *, ILO 4 at medium level, similar to problems on problem sheet]
$t^4 + 2t^2$ and t^3 are polynomials, elements of $\mathbb{C}[\mathbb{A}^1] = \mathbb{C}[t]$, so φ is morphism $\mathbb{A}^1 \to \mathbb{A}^2$. In order to prove that it is a morphism $\mathbb{A}^1 \to C$, we need to show that $\varphi(t) \in C$ for every $t \in \mathbb{C}$. If we substitute $x = t^4 + 2t^2$ and $y = t^3$ into f, we get

$$\begin{aligned}(t^4 + 2t^2)^3 - 6(t^4 + 2t^2)t^6 - t^{12} - 8t^6 &= t^6(t^2 + 2)^3 - t^6(6t^4 + 12t^2 + t^6 + 8)\\ &= t^6(t^2 + 2)^3 - t^6(t^2 + 2)^3 = 0.\end{aligned}$$

Therefore $\varphi(t) \in C$ for every $t \in \mathbb{C}$, so φ is indeed a morphism $\mathbb{A}^1 \to C$.

(ii) [4 marks, *, ILO 4 at medium level, similar to problems on problem sheet]
$xy+4y$ and $2x+y^2$ (strictly speaking, their cosets) are elements of $\mathbb{C}[C]$. $2x+y^2$ is not a multiple of f, for example, because its degree is smaller than the degree of f, therefore $2x + y^2$ is not identically 0 on C. Hence $\dfrac{xy + 4y}{2x + y^2}$ is a rational

function on C, so it gives a rational map $C \dashrightarrow \mathbb{A}^1$.

$$(\psi \circ \varphi)(t) = \psi(t^4 + 2t^2, t^3) = \frac{(t^4 + 2t^2)t^3 + 4t^3}{2(t^4 + 2t^2) + t^6} = \frac{t^7 + 2t^5 + 4t^3}{2t^4 + 4t^2 + t^6} = t$$

for every $t \in \mathbb{C}$ where the denominators are not 0, therefore $\psi \circ \varphi$ is indeed the identity map.

(iii) [6 marks, *, ILO 3 at high level, similar to problems on problem sheet]
The Jacobian matrix is

$$(f_x \quad f_y) = (3x^2 - 6y^2 \quad -12xy - 4y^3 - 16y) = (3(x^2 - 2y^2) \quad -4y(y^2 + 3x + 4))$$

The points of C where $\operatorname{rank} J = 0$ are the points in $\mathbb{C}^2$ where $f = f_x = f_y = 0$.
$-4y(y^2 + 3x + 4) = 0$ implies $y = 0$ or $y^2 + 3x + 4 = 0$.
If $y = 0$, then $x^2 - 2y^2 = 0$ implies that $x = 0$, too. $(0, 0) \in C$ and $\operatorname{rank} J_{(0,0)} = 0$.
If $y^2 + 3x + 4 = 0$, then $y^2 = -3x - 4$, by substituting $-3x - 4$ for y^2 in $x^2 - 2y^2 = 0$ we get $x^2 + 6x + 8 = 0$, which has solutions $x = -2$ and $x = -4$. If $x = -2$, then $y^2 = -3x - 4 = 2$, so $y = \pm\sqrt{2}$, but $f(-2, \pm\sqrt{2}) = -4$, so $(2, \pm\sqrt{2}) \notin C$. If $x = -4$, then $y^2 = -3x - 4 = 8$, so $y = \pm 2\sqrt{2}$. $f(-4, \pm 2\sqrt{2}) = 0$, so $(-4, \pm\sqrt{2}) \in C$ and $\operatorname{rank} J_{(-4, \pm 2\sqrt{2})} = 0$.
Therefore the points of C where $\operatorname{rank} J = 0$ are $(0, 0)$, $(-4, 2\sqrt{2})$ and $(-4, -2\sqrt{2})$. The rank of the Jacobian is either 0 or 1, therefore the rank is 1 at all other points of C and these 3 points are the singular points of C.

3. (a)(i) [1 mark, *, ILO 1 at low level]
The projective algebraic variety defined by I is the set

$$\mathcal{V}(I) = \{(X_0 : X_1 : \ldots : X_n) \in \mathbb{P}^n \,|F(X_0, X_1, \ldots, X_n) = 0 \text{ for every homogeneous } F \text{ in } I\}.$$

(ii) [2 marks, *, ILO 1 at medium level]
$\mathcal{I}(W) \lhd K[X_0, X_1, \ldots, X_n]$ is the ideal of $K[X_0, X_1, \ldots, X_n]$ generated by the set

$$\{F \in K[X_0, X_1, \ldots, X_n] \,|F \text{ is homogeneous}, F(X_0, X_1 \ldots, X_n) = 0 \;\forall (X_0 : \ldots : X_n) \in W\}.$$

$K[W] = K[X_0, X_1, \ldots, X_n]/\mathcal{I}(W)$.
(b)(i) [2 marks, *, ILO 2 at medium level, bookwork]
Let a_{ij}, $0 \le i \le n$, $0 \le j \le m$, be the entries of A and let $\Phi_i(X_0, X_1 \ldots, X_m) = \sum_{j=0}^{m} a_{ij} X_j$. Φ_0, Φ_1, $\ldots$, Φ_n are the components of Φ, they are all homogeneous polynomials of degree 1. As A is not the 0 matrix, there exists i, $0 \le i \le n$, such that $\Phi_i \ne 0$ (as an element of $K[\mathbb{P}^m] = K[X_0, X_1, \ldots, X_m]$), therefore Φ is a rational map $\mathbb{P}^m \dashrightarrow \mathbb{P}^n$.
(ii) [2 marks, *, ILO 2 at medium level, bookwork]
If A is invertible, then $A \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ implies $X_0 = X_1 = \ldots = X_m = 0$, so for any $(X_0 : X_1 : \ldots : X_m) \in \mathbb{P}^m$, $A \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_m \end{pmatrix} \ne \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, therefore Φ is defined at $(X_0 : X_1 : \ldots : X_m)$. As Φ is defined at every point of $\mathbb{P}^m$, it is a morphism.
(iii) [3 marks, ILO 4 at medium level, unseen, but not difficult]
The condition on the rows of A means that Φ_0, Φ_1, $\ldots$, Φ_n are coprime and since $K[\mathbb{P}^m] = K[X_0, X_1, \ldots, X_m]$ is a UFD, we only need to consider this representation of Φ to determine whether Φ is defined at a particular point as all other representations are of the form $(F\Phi_0 : F\Phi_1 : \ldots : F\Phi_n)$, where F is a homogeneous polynomial. As $m > n$, there exist a vector $\begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_m \end{pmatrix} \ne \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ such

that $A\begin{pmatrix}X_0\\X_1\\\vdots\\X_m\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}$. Then $\Phi_i(X_0, X_1 \ldots, X_m) = 0$ for every i, $0 \le i \le n$, therefore Φ is not defined at $(X_0 : X_1 : \ldots : X_n)$, so Φ is not a morphism.

(c) [5 marks, ILO 5 at high level, bookwork]

$\varphi(z) = \dfrac{az+b}{cz+d}$ for some a, b, c, $d \in K$, $ad - bc \neq 0$. If $c = 0$, $\varphi(z) = \dfrac{a}{d}z + \dfrac{b}{d}$, while if $c \neq 0$, $\varphi(z) = \dfrac{bc-ad}{c} \cdot \dfrac{1}{cz+d} + \dfrac{a}{c}$. In either case, φ can be composed of maps of the form $z \mapsto z + \alpha$ ($\alpha \in K$), $z \mapsto \lambda z$ ($\lambda \in K \setminus \{0\}$) and $z \mapsto 1/z$. It is easy to see that the first two preserve the cross ratio and

$$\begin{aligned}\left(\frac{1}{z_1}, \frac{1}{z_2}; \frac{1}{z_3}, \frac{1}{z_4}\right) &= \frac{\left(\frac{1}{z_1} - \frac{1}{z_3}\right)\left(\frac{1}{z_2} - \frac{1}{z_4}\right)}{\left(\frac{1}{z_1} - \frac{1}{z_4}\right)\left(\frac{1}{z_2} - \frac{1}{z_3}\right)} = \frac{\left(\frac{z_3 - z_1}{z_1 z_3}\right)\left(\frac{z_4 - z_2}{z_2 z_4}\right)}{\left(\frac{z_4 - z_1}{z_1 z_4}\right)\left(\frac{z_3 - z_2}{z_2 z_3}\right)}\\ &= \frac{(z_3 - z_1)(z_4 - z_2)}{(z_4 - z_1)(z_3 - z_2)} = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} = (z_1, z_2; z_3, z_4).\end{aligned}$$

All three types of maps preserve the cross ratio, therefore so does φ.

(d) [5 marks, ILO 5 at high level, similar to problems on problem sheet]

φ preserves the cross ratio, therefore $(-1, 3; 5, z) = (\varphi(-1), \varphi(3); \varphi(5), \varphi(z)) = (-5, 3; 4, \varphi(z))$ for any $z \in \mathbb{C}$. $(-1, 3; 5, z) = \dfrac{(-6)(3-z)}{(-1-z)(-2)} = \dfrac{3(z-3)}{z+1}$ and $(-5, 3; 4, \varphi(z)) = \dfrac{(-9)(3-\varphi(z))}{(-5-\varphi(z))(-1)} = \dfrac{9(\varphi(z)-3)}{\varphi(z)+5}$, so we obtain the equation

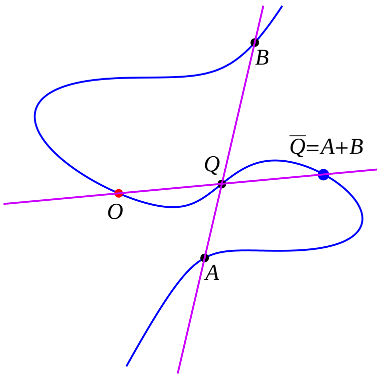
$$\frac{3(z-3)}{z+1} = \frac{9(\varphi(z)-3)}{\varphi(z)+5}.$$

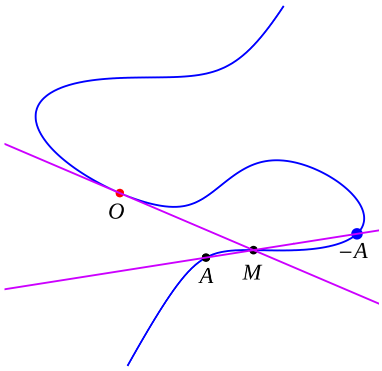
By multiplying both sides by $(z+1)(\varphi(z)+5)/3$, we get

$$\begin{aligned}(z-3)(\varphi(z)+5) &= 3(z+1)(\varphi(z)-3)\\ z\varphi(z) - 3\varphi(z) + 5z - 15 &= 3z\varphi(z) + 3\varphi(z) - 9z - 9\\ -2z\varphi(z) - 6\varphi(z)5 &= -14z + 6\\ \varphi(z) &= \frac{-14z+6}{-2z-6} = \frac{7z-3}{z+3}\end{aligned}$$

Therefore $a = 7$, $b = -3$, $c = 1$ and $d = 3$ is a solution. (The solution is not unique, a, b, c, d can be multiplied by a non-0 complex number to get another solution.)

4. (a)(i) [3 marks, *, ILO 6, low level]
For $A, B \in E$ define $A + B \in E$ as follows: let Q be the third intersection point of the line AB with E, and then $A + B$ is the third intersection point of the line OQ with E as shown on the diagram below on the left.
If $A = B$ or $O = Q$, then the line AB or the line OQ is taken to be the tangent line at A or O, resp. If any line in this construction is tangent to E, the third intersection point is defined using intersection multiplicities.





(ii) [4 marks, 2* for describing the contruction of $-A$, 2 for proving $A + (-A) = 0$, ILO 6 at low and medium level, bookwork]
The construction is shown on the diagram above on the right. Take the tangent line at O and let M be its 3rd point of intersection with E. To find $-A$, take the line through A and M, its 3rd point of intersection with E is $-A$. (The cases when two points coincide or when a line is tangent to E are dealt with as in (i).)
The first step in adding A and $-A$ is to find 3rd intersection point of the line through A and $-A$ with E. By the construction of $-A$, this 3rd intersection point is M. The 2nd step is to take the 3rd intersection point of the line OM with E, as this line is tangent to E at O, the 3rd intersection point is defined to be O, so $A + (-A) = O$ indeed.

(b) [9 marks, 4 for $A + B$, 5 for $2A$, *, ILO 7 at high level]
The equation of the line AB is $y = 6 - 3x$. If we substitute $6 - 3x$ for y in the equation of F, we get

$$\begin{aligned}(6-3x)^2 &= x^3 - 4x^2 + 8x + 4,\\ 36 - 36x + 9x^2 &= x^3 - 4x^2 + 8x + 4,\\ 0 &= x^3 - 13x^2 + 44x - 32.\end{aligned}$$

We know that $x = 1$ and $x = 4$ are roots of this cubic equation, so by the relation between the roots and the coefficients, the third root must be $x = 8$.

This is the x coordinate of the 3rd intersection point of the line AB with E, the y coordinate is $6 - 3 \times 8 = -18$, therefore $A + B = (8, 18)$.

To calculate $2A$, we first need to calculate the equation of tangent line at A. Let $f(x, y) = x^3 - 4x^2 + 8x + 4 - y^2$, so that the equation of F is $f(x, y) = 0$. Then $\dfrac{\partial f}{\partial x} = 3x^2 - 8x + 8$, $\dfrac{\partial f}{\partial y} = -2y$, their values at $A = (1, 3)$ are 3 and -6, resp., so the slope of the tangent line to F at A is $-\dfrac{3}{-6} = \dfrac{1}{2}$ and the equation of the tangent line is $y = \dfrac{1}{2}x + \dfrac{5}{2}$.

If we substitute $\dfrac{1}{2}x + \dfrac{5}{2}$ for y in the equation of F, we get

$$\begin{aligned}
\left(\frac{1}{2}x + \frac{5}{2}\right)^2 &= x^3 - 4x^2 + 8x + 4, \\
\frac{x^2}{4} + \frac{5}{2}x + \frac{25}{4} &= x^3 - 4x^2 + 8x + 4, \\
0 &= x^3 - \frac{17}{4}x^2 + \frac{11}{2}x - \frac{9}{4}.
\end{aligned}$$

We know that $x = 1$ is a double root of this cubic equation, so by the relation between the roots and the coefficients, the third root must be $x = \dfrac{9}{4}$. This is the x coordinate of the "3rd" intersection point of the tangent line to E at A with E, the y coordinate is $\dfrac{1}{2} \cdot \dfrac{9}{4} + \dfrac{5}{2} = \dfrac{29}{8}$, therefore $2A = \left(\dfrac{9}{4}, -\dfrac{29}{8}\right)$.

(b) [4 marks, ILO 7 at medium level, it is not difficult, but this is very last topic in the course and the students have to recognise that they do not have to do the whole procedure]

We carry out changes of variables corresponding to invertible affine transformations. The first substitution is $x_1 = 3x$, $y_1 = 3y$, this makes the coefficients of y_1^2 and x_1^3 equal, we obtain

$$y_1^2 - 2x_1y_1 - 4y_1 = x_1^3 - x_1^2.$$

We now complete the square with respect to to y_1, so $y_2 = y_1 - x_1 - 2$, $x_2 = x_1$. $y_1^2 - 2x_1y_1 - 4y_1 = y_2^2 - (x_1 + 2)^2 = y_2^2 - (x_2 + 2)^2$, thus we obtain

$$\begin{aligned}
y_2^2 &= x_2^3 - x_2^2 + (x_2 + 2)^2, \\
y_2^2 &= x_2^3 + 4x_2 + 4.
\end{aligned}$$

This equation is already in the required form, therefore $p = q = 4$ is a solution.

Solutions for MATH34032, Green's Functions, Integral Equations, and Applications, 2019 Final Exam.

Question	Learning Outcome	Solution
Q1(a)	Recognise regular Sturm-Liouville boundary value and eigenvalue problems ... . This question assesses the ILO at a low level.	This is not a regular Sturm-Liouville eigenvalue problem because the coefficient of the second derivative, $\cos(x)$ is equal to zero at the endpoints of the interval. Two marks for some argument that uses a valid property of regular S-L eigenvalue problems, one mark for the right answer, then two marks for the correct explanation. [5 marks]
Q1(b)(i)	Construct Green's functions for one dimensional boundary value problems from fundamental solutions, and use these Green's functions to express solutions to such problems. The adjoint Green's function was used in deriving the formula for solutions of BVPs with inhomogeneous boundary conditions. This question is asking students to recall a fact covered in lecture (bookwork), and assesses the ILO at a low level.	The adjoint Green's function is the Green's function for the adjoint problem. It satisfies $$\begin{cases} \mathcal{L}_X^* G^*(x,x_0) = \delta(x-x_0), \quad x,\ x_0 \in (a,b) \\ \mathcal{B}_x^*,\ x_0 \in (a,b) \end{cases}$$ One mark for knowing G^* is the Green's function of the adjoint problem, one mark for the BVP for G^* (both marks can be earned by writing the BVP as well). [2 marks]
Q1(b)(ii)	Construct Green's functions for one dimensional boundary value problems from fundamental solutions, and use these Green's functions to express solutions to such problems. The reciprocity relation was used in deriving the formula for solutions of BVPs with inhomogeneous boundary conditions. This question is asking students to recall a fact covered in lecture (bookwork), and assesses the ILO at a low level.	The adjoint Green's function and Green's function are related by reciprocity: $$G(x,x_0) = \overline{G^*(x_0,x)}.$$ One mark for knowing the term reciprocity, two for formula. All three marks can be earned by the formula alone. [3 marks]
Q1(c)	Prove that the Laplacian of the free-space Green's functions in two and three dimensions equals the Dirac delta function	The free space Green's function for the Laplacian in three dimensions is $$G(\mathbf{x},\mathbf{x}_0) = -\frac{1}{4\pi}\frac{1}{\|\mathbf{x}-\mathbf{x}_0\|}.$$

	This is a basic recall question. It assesses the ILO at a low level.	One mark will be awarded if the two dimensional Green's function is given instead. Minor errors in recalling the Green's function may be penalised one mark. [3 marks]
Q1(d)(i)	Distinguish between the different types of integral equations ... . This question assesses the ILO at a low level.	This is a Fredholm integral equation of the second kind. One mark for Fredholm, one mark for second kind. [2 marks]
Q1(d)(ii)	Distinguish between the different types of integral equations ... . This question assesses the ILO at a low level.	The kernel is $k(x, y) = e^{-(x-y)^2}$. This will be marked all or nothing. [2 marks]
Q1(e)	Apply the Fredholm alternative to one dimensional boundary value problems to determine whether solutions exist, and whether they are unique. This question assesses the ILO at a low level.	The condition is $$\langle v, f\rangle = 0$$ where the angle brackets indicate the $L^2[a, b]$ inner product. [6 marks]
Q2	Construct Green's functions for one dimensional boundary value problems from fundamental solutions ... . This question assesses the ILO at a medium level.	Since the boundary conditions are separated in this example we can use the general formula for the Green's function in the case of separated boundary conditions. This is $$G(x, x_0) = \frac{u_1(x)u_2(x_0)}{p(x_0)W(x_0)}H(x_0 - x) + \frac{u_1(x_0)u_2(x)}{p(x_0)W(x_0)}H(x - x_0) \quad (1)$$ where u_1 and u_2 form a complementary solution and $u_1(0) = 0$, $u_2(1) = 0$. We must find u_1 and u_2 satisfying these conditions. First, we can take $u_1(x) = xe^{-x^2}$. Second, we can take $u_2(x) = (x - 1)e^{-x^2}$. To apply (1) we must next calculate the Wronskian $$\begin{aligned} W(x) &= u_1(x)u_2'(x) - u_1'(x)u_2(x) \\ &= e^{-2x^2}x(1 - 2x(x - 1)) - e^{-2x^2}(1 - 2x^2)(x - 1) \\ &= e^{-2x^2}(x - 2x^3 + 2x^2 - x + 1 + 2x^3 - 2x^2) \\ &= e^{-2x^2}. \end{aligned}$$ Putting this into (1) we have $$G(x, x_0) = x(x_0 - 1)e^{x_0^2 - x^2}H(x_0 - x) + x_0(x - 1)e^{x_0^2 - x^2}H(x - x_0).$$ Four marks for the general formula in the case of separated boundary conditions, two marks each for u_1 and u_2, one mark for the general formula for the Wronskian, two marks for correct Wronskian, one mark for final formula. [12 marks]

Q3(a)	Apply the Fredholm alternative to one dimensional boundary value problems ... . Finding the adjoint problem is part of applying the Fredholm alternative. We did show that Sturm-Liouville operators are formally self adjoint in lecture, and showing such problems are fully self adjoint is in the notes. Students have not been presented with a problem formulated exactly like this one however. This question assesses the ILO at a medium level.	We must calculate the adjoint operator and boundary conditions using integration by parts

$$\begin{aligned}\langle v, pu'' + p'u' \rangle_{L^2[0,1]} &= \int_0^1 p\overline{v}u'' + p'\overline{v}u' \; \mathrm{d}x \\ &= \int_0^1 (p\overline{v})''u - (p'\overline{v})'u \; \mathrm{d}x + \left(p\overline{v}u' - p\overline{v}'u + p'\overline{v}u\right|_0^1.\end{aligned}$$

Applying the boundary condition $u(0) = u(1) = 0$, and expanding the derivatives using the product rule we obtain

$$\langle v, pu'' + p'u' \rangle_{L^2[0,1]} = \int_0^1 \overline{(pv'' + p'v')}u \; \mathrm{d}x + p(1)\overline{v}(1)u'(1) - p(0)\overline{v}(0)u'(0).$$

From this we see that $\mathcal{L}^* v = pv'' + p'v'$ and so the operator is formally self adjoint. For the boundary conditions, since $u'(1)$ and $u'(0)$ are arbitrary and $p(1)$, $p(0) > 0$, we must have

$$\mathcal{B}^* = \{v(1) = v(0) = 0\}$$

for the boundary terms to vanish. Therefore the problem is fully self-adjoint.

Two marks for attempting integration by parts in a way that demonstrates understanding of what is required, Two marks for successfully integrating by parts, one mark each for conclusions about adjoint operator and boundary conditions. It is alright if they rewrite the ODE in a Sturm-Liouville form, which may make the integration by parts easier.

[6 marks]

Q3(b)	Apply the Fredholm alternative to one dimensional boundary value problems to determine whether solutions exist, and whether they are unique. This problem was covered in lecture, and so is bookwork. If it were not bookwork this would be a high level problem, but given that it is bookwork it is assessing the ILO at a medium level.	Since the problem is fully self adjoint by part (a), the homogeneous adjoint problem is (after rewriting the ODE slightly)

$$\begin{cases} (pv')'(x) = 0, & x \in (0,1) \\ v(0) = 0, \; v(1) = 0. \end{cases}$$

To show that there is a unique solution of the original problem we must show that the only solution of the homogeneous adjoint problem is the trivial solution $v(x) = 0$. To do this we consider, using the ODE satisfied by v,

$$0 = \int_0^1 (pv')'\overline{v} \; \mathrm{d}x. \tag{2}$$

Integrating by parts and using the boundary condition on v gives

$$0 = -\int_0^1 p|v'|^2 \; \mathrm{d}x.$$

Since $p(x) > 0$, this implies that $v'(x) = 0$ for all x, and so v must be constant. By the boundary condition then $v(x) = 0$, and so the Fredholm alternative says that there is a unique solution. Two marks for knowing what that it is necessary to prove the only solution of the homogeneous adjoint problem is $v(x) = 0$, and one mark for writing down the homogeneous adjoint problem. One mark for (2), one mark for integrating by parts, and one mark for completing argument.

[6 marks]

Q4(a) ... solve Fredholm integral equations of the second kind with degenerate kernel. Students will have seen similar problems, but not this exact one. This question assesses the ILO at a low level.

To solve the equation first define

$$c = \int_0^1 y^2 u(y)\ \mathrm{d}y.$$

Then the integral equation becomes

$$u(x) = \lambda c x^3 + f(x).$$

Now multiply this by x^2 and integrate to obtain

$$\begin{aligned} c &= \lambda c \int_0^1 x^5\ \mathrm{d}x + \int_0^1 x^2 f(x)\ \mathrm{d}x \\ &= \frac{\lambda c}{6} + \int_0^1 x^2 f(x)\ \mathrm{d}x. \end{aligned}$$

This is equivalent to

$$\left(1 - \frac{\lambda}{6}\right) c = \int_0^1 x^2 f(x)\ \mathrm{d}x.$$

From this we can see that there will only be a unique solution if $\lambda \neq 6$. Some of the marking on this problem will be split with the other parts of Q4 since some of the required steps are common for all of the various parts. One mark will be awarded for defining c, two marks will be given for multiplying equation by x^2 and integrating, one mark will be given for correctly evaluating the integral, and one mark will be given for the correct final solution ($\lambda \neq 6$).

[5 marks]

Q4(b) ... solve Fredholm integral equations of the second kind with degenerate kernel. Students will have seen similar problems, but not this exact one. This question assesses the ILO at a low level.

Continuing on from the previous solution, when $\lambda \neq 6$ we can solve for c:

$$c = \frac{6}{6-\lambda} \int_0^1 y^2 f(y)\ \mathrm{d}y.$$

Based on this we can find $u(x)$:

$$u(x) = \frac{6\lambda}{6-\lambda} \int_0^1 x^3 y^2 f(y)\ \mathrm{d}y + f(x).$$

One mark will be awarded for defining c, one mark will be given for multiplying the equation by y^2 and integrating, one mark will be given for finding c, and two marks for the correct final solution for u.

[5 marks]

Q4(c) ... solve Fredholm integral equations of the second kind with degenerate kernel. Students will have seen similar problems, but not this exact one. This question assesses the ILO at a low level.

If $\lambda = 6$, then final equation in the solution of Q4(a) gives

$$0 = \int_0^1 x^2 f(x)\ \mathrm{d}x.$$

If there is a solution of the integral equation, then f must satisfy this condition. If f satisfies this condition then

$$u(x) = cx^3 + f(x)$$

is a solution of the integral equation for any constant c. Three marks will be given for the condition on f, and three marks for the solution for u.

[6 marks]

Q5(a) Formulate boundary value problems modelling waves on a string ... . This problem is bookwork. It assesses the ILO at a medium level.

The radiation conditions are

$$\lim_{x\to\infty} u'(x) - ik(x)u(x) = 0, \qquad \lim_{x\to-\infty} u'(x) + ik(x)u(x) = 0. \tag{3}$$

Suppose that $C > 0$ is sufficiently large so that $f(x) = 0$ and $k(x) = k_0$ when $x > C$. Then on the interval $x > C$

$$u''(x) + k(x)^2 u(x) = 0 \Leftrightarrow u(x) = b_1 e^{ik_0 x} + b_2 e^{-ik_0 x}.$$

The boundary condition as $x \to \infty$ is thus equivalent to

$$u'(x) - ik_0 u(x) = -2ik_0 b_2 e^{-ik_0 x} = 0 \quad \text{for } x > C \Leftrightarrow b_2 = 0.$$

Therefore the time-dependent waves on $x > C$ are given by the real part of

$$U(x) = b_1 e^{i(k_0 x - \omega t)},$$

which are waves moving to the right. Therefore all waves sufficiently far to the right are moving to the right, which means no waves are coming from positive infinity.

On the other hand, suppose that $C' < 0$ is sufficiently negative so that $f(x) = 0$ and $k(x) = k_0$ when $x < C'$. Then on the interval $x < C'$

$$u''(x) + k(x)^2 u(x) = 0 \Leftrightarrow u(x) = a_1 e^{ik_0 x} + a_2 e^{-ik_0 x}.$$

The boundary condition $x \to -\infty$ is thus equivalent to

$$u'(x) + ik_0 u(x) = 2ik_0 a_1 e^{ik_0 x} = 0 \quad \text{for } x < C' \Leftrightarrow a_1 = 0.$$

Therefore the time-dependent waves on $x < C'$ are given by the real part of

$$U(x) = a_2 e^{-i(k_0 x + \omega t)},$$

which are waves moving to the left. Therefore all waves sufficiently far to the left are moving to the left, which means no waves are coming from negative infinity.

One mark each for the two conditions, one mark for formulas for u when $|x|$ is sufficiently large, two marks for applying boundary conditions, one mark for connecting this to the time dependent waves.

[6 marks]

Q5(b) Formulate boundary value problems modelling waves on a string, and apply the Neumann

We have

$$u''(x) + k(x)^2 u(x) = 0.$$

	series solution for a corresponding Fredholm integral equation of the second kind to analyze scattering of waves in one dimension. This problem is bookwork. It assesses the ILO at a medium level.	Putting in $u(x) = e^{ik_0x} + u_{sc}(x)$ we have $$u''_{sc}(x) + k(x)^2 u_{sc}(x) = (k_0^2 - k(x)^2)e^{ik_0x} \Rightarrow u''_{sc}(x) + k_0^2 u_{sc}(x) = (k_0^2 - k(x)^2)u(x).$$ Since u_{sc} also satisfies the radiation conditions we can apply the Green's function given in the problem to get $$u_{sc}(x) = \int_{-\infty}^{\infty} e^{ik_0\|x-x_0\|}\frac{k_0^2 - k(x_0)^2}{2ik_0}u(x_0)\ \mathrm{d}x_0.$$ Adding e^{ik_0x} we get that u satisfies the Fredolm integral equation $$u(x) = e^{ik_0x} + \int_{-\infty}^{\infty} e^{ik_0\|x-x_0\|}\frac{k_0^2 - k(x_0)^2}{2ik_0}u(x_0)\ \mathrm{d}x_0.$$ This problem is bookwork. One mark for putting form for u into ODE, two marks for arriving at non-homogeneous equation for u_{sc}, two marks for applying the Green's function, one mark for adding e^{ik_0x} to get the final equation for u. [6 marks]
Q5(c)	Formulate boundary value problems modelling waves on a string, and apply the Neumann series solution for a corresponding Fredholm integral equation of the second kind to analyze scattering of waves in one dimension. This problem is bookwork. Assuming students can do the previous part it assesses the ILO at a low level, but the previous part is medium level.	The Born approximation is the first two terms in the Neumann series solution for the integral equation. In this case this gives $$u(x) \approx e^{ik_0x} + \int_{-\infty}^{\infty} e^{ik_0\|x-x_0\|}\frac{k_0^2 - k(x_0)^2}{2ik_0}e^{ik_0x_0}\ \mathrm{d}x_0.$$ Two marks for knowing what the Born approximation is (first two terms of Neumann series), and two marks for the formula in this case. [4 marks]
Q5(d)	Formulate boundary value problems modelling waves on a string, and apply the Neumann series solution for a corresponding Fredholm integral equation of the second kind to analyze scattering of waves in one dimension. Students have not see this problem, although they will have seen solutions of the scattering problem in some other cases, and	If $k(x)^2 = 2 + \delta(x)$, then $k_0^2 = 2$, and $k_0^2 - k(x)^2 = -\delta(x)$. Therefore the integral equation from part (c) becomes $$\begin{aligned} u(x) &= e^{2ix} - \frac{1}{4i}\int_{-\infty}^{\infty} e^{2i\|x-x_0\|}\delta(x_0)u(x_0)\ \mathrm{d}x_0 \\ &= e^{2ix} + \frac{e^{2i\|x\|}}{4i}u(0). \end{aligned} \quad (4)$$ If we evaluate the last equation at $x = 0$ we get $$u(0) = 1 - \frac{1}{4i}u(0) \Rightarrow u(0) = \frac{4i}{4i+1}.$$

there is a similar problem on a previous final exam which they will be able to access. It is assessing the ILO at a high level.

Putting this back into (4) we find that

$$u(x) = e^{2ix} + \frac{e^{2i|x|}}{4i+1}.$$

[6 marks]

Q6(a)

Prove that the Laplacian of the free-space Green's functions in two and three dimensions equals the Dirac delta function ... . Students will have seen the same calculation for the Green's function Laplacian, but this is new and may be difficult. It is assessing the ILO at a high level.

For $\mathbf{x} \neq \mathbf{x_0}$ we first of all have that the gradient of G is given by

$$\nabla_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) = -\frac{e^{ik|\mathbf{x}-\mathbf{x}_0|}}{4\pi} \left(\frac{ik}{|\mathbf{x}-\mathbf{x}_0|^2} - \frac{1}{|\mathbf{x}-\mathbf{x}_0|^3} \right) (\mathbf{x}-\mathbf{x}_0).$$

Taking the divergence of this we get

$$\begin{aligned}
\nabla^2_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) &= -\frac{ike^{ik|\mathbf{x}-\mathbf{x}_0|}}{4\pi} \left(\frac{ik}{|\mathbf{x}-\mathbf{x}_0|} - \frac{1}{|\mathbf{x}-\mathbf{x}_0|^2} \right) + \frac{e^{ik|\mathbf{x}-\mathbf{x}_0|}}{4\pi} \left(\frac{2ik}{|\mathbf{x}-\mathbf{x}_0|^2} - \frac{3}{|\mathbf{x}-\mathbf{x}_0|^3} \right) \\
&\quad - \frac{3e^{ik|\mathbf{x}-\mathbf{x}_0|}}{4\pi} \left(\frac{ik}{|\mathbf{x}-\mathbf{x}_0|^2} - \frac{1}{|\mathbf{x}-\mathbf{x}_0|^3} \right) \\
&= k^2 \frac{1}{4\pi} \frac{e^{ik|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}.
\end{aligned}$$

Therefore

$$\nabla^2_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) + k^2 G(\mathbf{x}, \mathbf{x}_0) = 0.$$

The steps in the calculation need not be the same as given here. Generally a minimum of three marks will be awarded for students who make a reasonable effort at the calculation but do not succeed. Three marks will be given for computing the gradient, and then three marks for computing the Laplacian.

[9 marks]

Q6(b)

Prove that the Laplacian of the free-space Green's functions in two and three dimensions equals the Dirac delta function ... . The equivalent calculation for the Laplacian in three dimensions is covered in lecture, but this is new. It is assessing the ILO at a high level.

We apply Green's formula from the cover of the exam with $D = \mathbb{R}^3 \setminus B_\epsilon(\mathbf{x}_0)$, and $g(\mathbf{x}) = G(\mathbf{x}, \mathbf{x}_0)$ to get

$$\begin{aligned}
&\int_{\mathbb{R}^3 \setminus B_\epsilon(\mathbf{x}_0)} (f(\mathbf{x}) \nabla^2_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) - G(\mathbf{x}, \mathbf{x}_0) \nabla^2 f(\mathbf{x}))\ d\mathbf{x} \\
&\qquad = \int_{\partial B_\epsilon(\mathbf{x}_0)} (f(\mathbf{x}) \nabla_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) - G(\mathbf{x}, \mathbf{x}_0) \nabla f(\mathbf{x})) \cdot \mathbf{n}(\mathbf{x})\ ds(\mathbf{x})
\end{aligned}$$

where $\mathbf{n}(\mathbf{x}) = (\mathbf{x}_0 - \mathbf{x})/\epsilon$ is the unit normal vector pointing into $B_\epsilon(\mathbf{x}_0)$. Note that this is allowed, and there is no contribution from any other boundary since $f(x) = 0$ for $|x|$ sufficiently large. Next we add and subtract $k^2 f(\mathbf{x}) G(\mathbf{x}, \mathbf{x}_0)$ in the integral on the left hand side to get

$$\begin{aligned}
&\int_{\mathbb{R}^3 \setminus B_\epsilon(\mathbf{x}_0)} (f(\mathbf{x})(\nabla^2_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) + k^2 G(\mathbf{x}, \mathbf{x}_0)) - G(\mathbf{x}, \mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ d\mathbf{x} \\
&\qquad = \int_{\partial B_\epsilon(\mathbf{x}_0)} (f(\mathbf{x}) \nabla_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}_0) - G(\mathbf{x}, \mathbf{x}_0) \nabla f(\mathbf{x})) \cdot \mathbf{n}(\mathbf{x})\ ds(\mathbf{x}).
\end{aligned}$$

Using part (a) the first term in the integrand on the left is zero, and so

$$\int_{\mathbb{R}^3\setminus B_\epsilon(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ d\mathbf{x} = -\int_{\partial B_\epsilon(\mathbf{x}_0)} (f(\mathbf{x})\nabla_{\mathbf{x}} G(\mathbf{x},\mathbf{x}_0) - G(\mathbf{x},\mathbf{x}_0)\nabla f(\mathbf{x}))\cdot \mathbf{n}(\mathbf{x})\ ds(\mathbf{x}).$$

Next, using the formula we found in part (a) for $\nabla_{\mathbf{x}} G(\mathbf{x},\mathbf{x}_0)$ we have

$$\begin{aligned}\int_{\mathbb{R}^3\setminus B_\epsilon(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ d\mathbf{x} = -\frac{1}{4\pi\epsilon}\int_{\partial B_\epsilon(\mathbf{x}_0)} e^{ik|\mathbf{x}-\mathbf{x}_0|}\Bigg(&\frac{f(\mathbf{x})}{|\mathbf{x}-\mathbf{x}_0|^3}(\mathbf{x}-\mathbf{x}_0) \\ &- \frac{ikf(\mathbf{x})}{|\mathbf{x}-\mathbf{x}_0|^2}(\mathbf{x}-\mathbf{x}_0) + \frac{\nabla f(\mathbf{x})}{|\mathbf{x}-\mathbf{x}_0|}\Bigg)\cdot(\mathbf{x}_0 - \mathbf{x})\ ds(\mathbf{x}).\end{aligned}$$

Finally, using the fact that on $B_\epsilon(\mathbf{x}_0)$, $|\mathbf{x}-\mathbf{x}_0| = \epsilon$, gives

$$\int_{\mathbb{R}^3\setminus B_\epsilon(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ d\mathbf{x} = \frac{e^{ik\epsilon}}{4\pi\epsilon^2}\int_{\partial B_\epsilon(\mathbf{x}_0)} (1-ik\epsilon)f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot(\mathbf{x}-\mathbf{x}_0)\ ds(\mathbf{x}).$$

Four marks for using Green's formula, one mark for correct identification of the normal vector, one mark for the use of formula for $\nabla_{\mathbf{x}} G$, one mark for use of fact $|\mathbf{x}-\mathbf{x}_0| = \epsilon$ on ball, one mark for final answer.

[8 marks]

Q6(c)

Prove that the Laplacian of the free-space Green's functions in two and three dimensions equals the Dirac delta function ... . As in the other parts, the version of this calculation for the Laplacian is covered in lecture. It is assessing the ILO at a high level.

First we make the change of variables suggested, $\mathbf{r} = (\mathbf{x}-\mathbf{x}_0)/\epsilon$ on the right hand side of the formula from (b) to get

$$\int_{\mathbb{R}^3\setminus B_\epsilon(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ \mathrm{d}\mathbf{x} = \frac{e^{ik\epsilon}}{4\pi}\int_{\mathbb{S}^2} (1-ik\epsilon)f(\mathbf{x}_0+\epsilon\mathbf{r}) + \epsilon\nabla f(\mathbf{x}_0+\epsilon\mathbf{r})\cdot\mathbf{r}\ ds(\mathbf{r})$$

Now, since f has continuous derivatives, $|\nabla f|$ is bounded, and so upon taking the limit the second term on the right goes to zero, and for the other terms we get in the limit $\epsilon \to 0^+$

$$\int_{\mathbb{R}^3} G(\mathbf{x},\mathbf{x}_0)(\nabla^2 f(\mathbf{x}) + k^2 f(\mathbf{x}))\ \mathrm{d}\mathbf{x} = f(\mathbf{x}_0)\frac{1}{4\pi}\int_{\mathbb{S}^2}\ \mathrm{d}r = f(\mathbf{x}_0),$$

where we have used the fact that the area of the unit sphere is 4π as given on the front of the exam.

Three marks for making the change of variables correctly, one mark for why ∇f term goes to zero, and one mark for using the volume of the sphere.

[5 marks]

MATH35012: Wave Motion.

ILO material as required by the school:

Intended Learning Outcomes

1. Define the basic properties of a wave.
2. Describe and classify the physical properties of a wave from its mathematical form.
3. Derive the dispersion relation for a range of wave problems.
4. Analyze the dispersion relation to draw physical conclusions.
5. Formulate a mathematical problem for a physically-described system, including (but not restricted to) the examples of elastic, water and sound waves.
6. Apply the methods of the course to previously unseen wave problems and variations of seen problems.

Levels: (L)ow, (M)edium, (H)igh

Q1. A variation of seen material to give a different dispersion relation. This is a straightforward question aimed at the tail end of the class.
ILOs: 1L,3L.

Q2. Entirely unseen. A more challenging 2D example requiring a deeper understanding.
ILOs: 1M,2H,3M,4H.

Q3. Bookwork. A standard derivation covered in the lecture material.
ILOs: 1M,2M.

Q4. Previously seen material.
ILOs: 5M.

Q5. A variation of the standard problem discussed in the lecture notes, here extended to include the bending stiffness/rigidity of the elastic sheet.
ILOs: 3H,4H,5H,6H.

Q6. A variation of an examples sheet problem, here extended to a change in pipe cross sectional area.
ILOs: 2H,5M,6H.

Q1.

$$y_t + a y_x - b y_{xxx} = 0 \qquad a, b > 0.$$

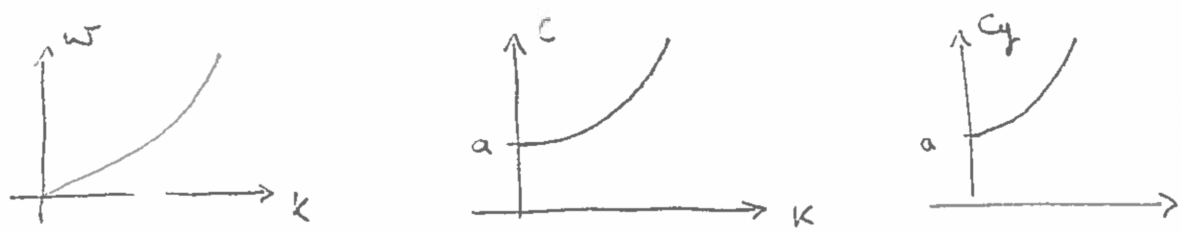
(i) $(-i\omega) + a(ik) - b(ik)^3 = 0.$

$$-\omega + ak + bk^3 = 0$$

$$\omega = ak + bk^3. \qquad [2]$$

(ii) $c = \omega/k$ phase speed & $c_g = \dfrac{d\omega}{dk}$ group vely.

$c = a + bk^2$ $\qquad$ $c_g = a + 3bk^2.$ $\qquad$ [2]

(iii) 

[3]

It is dispersive if $b \neq 0$.

(iv) $c_g > c$ and energy propagates faster than phase. [1]

8

Q2. $f_t - \beta R^2 f_x = R^2 \nabla^2 f_t \qquad R, \beta > 0.$

(i) $-i\omega - \beta R^2 (ik) = R^2(-i\omega)[-k^2 - l^2].$

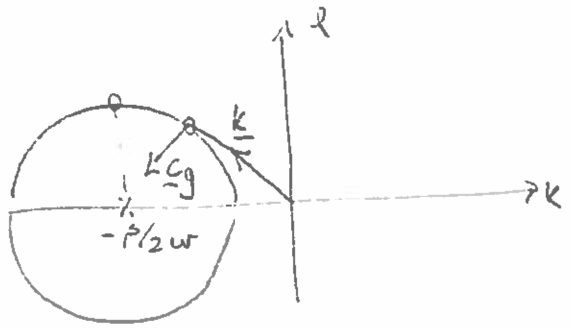
$$\omega + \beta R^2 k = R^2 \omega (-1)[k^2 + l^2]$$

$$k^2 R^2 \omega + \beta R^2 k + \omega + R^2 \omega l^2 = 0.$$

$$k^2 + \frac{\beta}{\omega} k + \frac{1}{R^2} + l^2 = 0.$$

$$\left(k + \frac{\beta}{2\omega}\right)^2 + l^2 = \frac{\beta^2}{4\omega^2} - \frac{1}{R^2}. \qquad [3]$$

(ii) for fixed $\omega > 0$, if $\frac{1}{R^2} < \frac{\beta^2}{4\omega^2}$ then we have circles with centre at $k = -\beta/2\omega$, $l = 0$ $\left(\omega_c < \frac{\beta^2}{4R^2}\right)$



$\underline{k} = \begin{pmatrix} k \\ l \end{pmatrix}$ wave no. vector

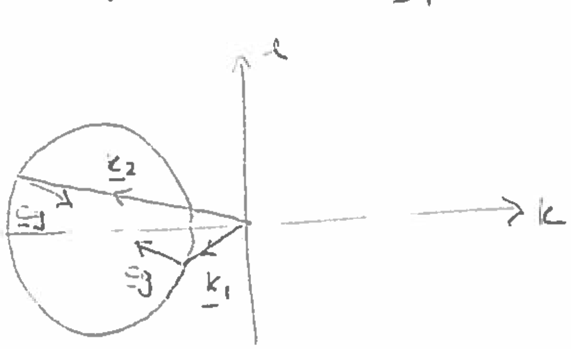
radius of circle is $\left(\frac{\beta^2}{4\omega^2} - \frac{1}{R^2}\right)^{1/2}$

$\underline{c}_g = \nabla_k \omega$ which is clearly perpendicular to $\omega = $ constant. [5]

[To see that $\underline{c}_g$ points inwards pick a value eg. $k = -\beta/2\omega$ & evaluate $\frac{\partial \omega}{\partial l}$.]

Too much algebra? → extra added.

(iii) wavelength $\lambda = 2\pi/|\underline{k}|$



$\underline{k}_1$: longer wave length

$\underline{k}_2$: shorter wave length

$\underline{k}_1$: energy propagates in -ve x direction. [4]

$\underline{k}_2$: energy propagates in +ve x direction.

Q2. additional requested by internal reviewer:

at $k = -\beta/2\omega$: $$2l = \frac{\beta^2}{4}(-2)\omega^{-3}\frac{\partial\omega}{\partial l}$$

so $\frac{\partial\omega}{\partial l} < 0$ if $l > 0$ and > 0 if $l < 0$

hence $\underline{c_g}$ points inwards.

Q3. [seen]

$$\phi = C\cosh(\hat{k}(z-h))\cos(\hat{k}x-\omega t) \qquad C \text{ const.}$$

where $\hat{k}\tanh\hat{k}h = \omega^2/g$.

For a particle at $\underline{r} = \begin{pmatrix} x \\ z \end{pmatrix}$ we know $\dfrac{d\underline{r}}{dt} = \underline{u} = \nabla\phi$ (1)

So
$$\left.\begin{aligned} \frac{dx}{dt} &= \phi_x = C\cosh(\hat{k}(z-h))(-\hat{k})\sin(\hat{k}x-\omega t) \\ \frac{dz}{dt} &= \phi_z = C\hat{k}\sinh(\hat{k}(z-h))\cos(\hat{k}x-\omega t) \end{aligned}\right\} - (*)$$ (2)

For small motion we can expand using a Taylor series:

eg. $\dfrac{\partial\phi}{\partial x} = \left.\dfrac{\partial\phi}{\partial x}\right|_{x_0,z_0} + $ smaller terms, $\dfrac{\partial\phi}{\partial z} = \left.\dfrac{\partial\phi}{\partial z}\right|_{z_0,x_0} +$ smaller

where (x_0, z_0) is a mean location. (1)

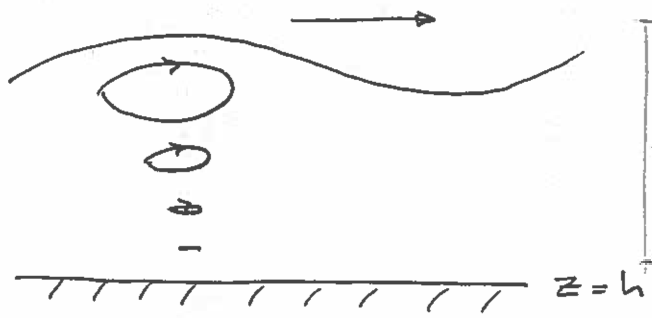
We can therefore integrate $(*)$ above:

$$x(t) = x_0 - \frac{\hat{k}C}{\omega}\cosh(\hat{k}(z-h))\cos(\hat{k}x-\omega t)$$

$$z(t) = z_0 - \frac{\hat{k}C}{\omega}\sinh(\hat{k}(z-h))\sin(\hat{k}x-\omega t)$$ (2)

$$\Rightarrow \frac{(x(t)-x_0)^2}{\cosh^2(\cdot)} + \frac{(z(t)-z_0)^2}{\sinh^2(\cdot)} = \frac{\hat{k}^2C^2}{\omega^2}$$ (2)

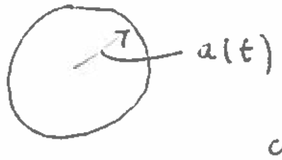
equation of an ellipse



(2)

10

Q4.

(i) 

boundary condition is impermeability of the sphere:

$$\frac{\partial \varphi}{\partial r} = \dot{a}(t) \quad \text{on } r = a(t). \qquad [2]$$

(ii) Either show by direct substitution or

$$\varphi_{tt} = \frac{c^2}{r^2}\frac{\partial}{\partial r}\left(r^2 \varphi_r\right) = \frac{c^2}{r^2}\left(2r\varphi_r + r^2\varphi_{rr}\right)$$

$$= \frac{c^2}{r}\left(2\varphi_r + r\varphi_{rr}\right) = \frac{c^2}{r}\frac{\partial^2}{\partial r^2}(r\varphi)$$

$[4]$

so $(r\varphi)_{tt} = c^2 (r\varphi)_{rr}$ the 1-D wave eq^{n}.

d'Alembert's solution: $r\varphi = f(t - r/c) + g(t + r/c)$

$g = 0$ (no incoming wave) & rewrite $f = -\frac{1}{4\pi} m$.

(iii) Using (i) $\varphi_r = -\frac{1}{4\pi}\left\{-r^{-2} m(t - r/c) + r^{-1} m'(t - r/c)\left(-\tfrac{1}{c}\right)\right\}$

$= \dot{a}(t)$ on $r = a(t)$.

$[4]$

$$\Rightarrow \frac{1}{4\pi}\left\{\frac{1}{a^2} m(t - a/c) + \frac{1}{ac} m'(t - a/c)\right\} = \dot{a}(t)$$

$\hat{t} = t - a_0/c$

$$m(\hat{t}) + \frac{a_0}{c} m'(\hat{t}) = \frac{d}{d\hat{t}} V(\hat{t} + a_0/c)$$

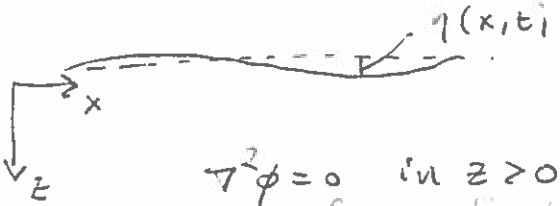
using $|a(t) - a_0| \ll 1$.

10

Q5. $\eta_t = \phi_z \qquad -\rho\phi_t = -\rho g \eta + T\eta_{xx} \mp R\,\eta_{xxxx}$ on $z=0$.

$\rho, g, T, R > 0$.

(i)



$\nabla^2\phi = 0$ in $z>0$ ϕ bounded as $z \to \infty$.

why no time derivative? — there is no $\partial/\partial t$.

and eliminating η: $\quad -\rho\,\phi_{tt} = -\rho g \eta_t + T\eta_{txx} \mp R\,\eta_{txxxx}$

then $\eta_t = \phi_z$, $\quad -\rho\phi_{tt} = -\rho g \phi_z + T\phi_{zxx} \mp R\,\phi_{zxxxx}$. [4]

(ii) $\phi(x,z,t) = \hat{X}(x)\hat{Z}(z)\hat{T}(t)$ $\quad \nabla^2\phi = 0 \Rightarrow \hat{X}''\hat{Z}\hat{T} + \hat{X}\hat{Z}''\hat{T} = 0$.

$$\frac{\hat{X}''}{\hat{X}} = -k^2 \qquad \frac{\hat{Z}''}{\hat{Z}} = +k^2 \qquad \text{for } k \in \mathbb{R}.$$

$$\hat{X} = A e^{ikx} + B e^{-ikx}, \qquad \hat{Z} = C e^{-kz} + D e^{+kz}.$$

For bounded/decaying sol$^{\underline{ns}}$ as $z\to\infty$ $D=0$ assuming (W.L.O.G) that $k>0$. [4]

$$\text{So } \phi = \left\{ \bar{A} e^{-kz} e^{ikx} + \bar{B} e^{-kz} e^{-ikx} \right\} \hat{T}(t)$$

Similarly from the surface condition $\dfrac{\hat{T}''}{\hat{T}} = -\omega^2$

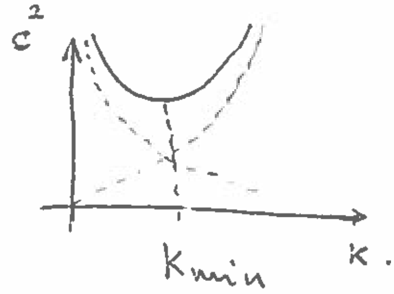
where $\quad -\rho(-\omega^2) = -\rho g(-k) + T(-k)(ik)^2 + R(\pm k)(ik)^4$

$$\omega^2 = gk + \frac{Tk^3}{\rho} + \frac{Rk^5}{\rho}. \qquad [2]$$

$\phi = \bar{A} e^{-kz} e^{i(kx \pm \omega t + \Psi)}$ for some phase shift Ψ.

Q5.

(iii) $k>0$ $\quad c^2 = \dfrac{\omega^2}{k^2} = g/k + \dfrac{Tk}{\rho} + \dfrac{Rk^3}{\rho}$.



$$2c\frac{dc}{dk} = 0 \Rightarrow -\frac{g}{k^2} + \frac{T}{\rho} + \frac{3Rk^2}{\rho} = 0$$

$$\frac{3Rk^4}{\rho} + \frac{Tk^2}{\rho} - g = 0 \quad \text{or} \quad k^4 + \frac{T}{3R}k^2 - \frac{\rho g}{3R} = 0.$$

$$k^2 = \left[-\frac{T}{3R} \pm \sqrt{\left\{\frac{T^2}{9R^2} + \frac{4\rho g}{3R}\right\}}\right]\frac{1}{2}$$

[5]

$$k_{min} = -\frac{T}{6R} + \sqrt{\left\{\frac{T^2}{(6R)^2} + \frac{\rho g}{3R}\right\}}$$

(clearly a minimum from the sketch).

(iv) $$\frac{c_g}{c} = \frac{d\omega/dk}{\omega/k} = \frac{2\omega\, d\omega/dk}{2\omega^2/k} = \frac{g + \frac{3Tk^2}{\rho} + \frac{5Rk^4}{\rho}}{2\left(g + \frac{Tk^2}{\rho} + \frac{Rk^4}{\rho}\right)}$$

$$= 1 + \frac{1}{2}\left\{\frac{-g + \frac{Tk^2}{\rho} + \frac{3Rk^4}{\rho}}{g + \frac{Tk^2}{\rho} + \frac{Rk^4}{\rho}}\right\}$$

[5]

$k = k_{min} \Rightarrow c_g = c$ non-dispersive

$k \to 0 \Rightarrow c_g = \frac{1}{2}c$ energy prop. slower than phase

$k \to \infty \Rightarrow c_g = \frac{3c}{2}$ —— " —— faster —— " —— .

(10)

35012

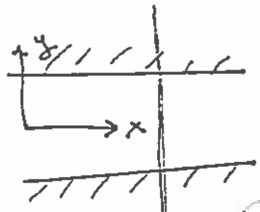
Q5. [seen similar]. changes in gas property not changes in cross-section.

(i) $\phi_{tt} = c^2 \nabla^2 \phi$, the wave equation. —— (1)

$\frac{\partial \phi}{\partial n} = 0$ at impermeable walls. —— (2)

2

$\phi = B\, e^{i\omega(t - x/c)}$



solves (1) $-\omega^2 = c^2 \cdot \left(\frac{-i\omega}{c}\right)^2 = -\omega^2$.

ϕ is indep. of y so $\frac{\partial \phi}{\partial n} = 0$ (everywhere).

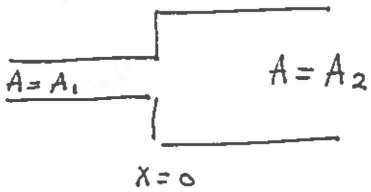
2

needed? - given in (iii) [not needed]

(ii) $\left\{ \begin{array}{l} \cancel{p = -\rho_0 \phi_t} \\ Q = \rho_0 A u = \rho_0 A \phi_x . \end{array} \right.$

2

(iii)



$\phi = \phi_I + \phi_R \qquad x < 0$

$\phi = \phi_T \qquad x > 0 .$

3

$\phi_I = B_I\, e^{i\omega(t - x/c)}$, $\phi_R = B_R\, e^{i\omega(t + x/c)}$, $\phi_T = B_T\, e^{i\omega(t - x/c)}$.

Continuity of $p \Rightarrow \left[\phi_t \right]_{x=0} = 0$ continuity of $Q \Rightarrow \left[A \phi_x \right]_{x=0} = 0$.

$i\omega B_I + i\omega B_R = i\omega B_T . \Rightarrow B_I + B_R = B_T .$ —— (3)

$A_1 \left[\left(-\frac{i\omega}{c}\right) B_I + \left(+\frac{i\omega}{c}\right) B_R \right] = A_2 \left(-\frac{i\omega}{c}\right) B_T$

2

$\Rightarrow A_1 B_R - A_1 B_I = -A_2 B_T .$ —— (4)

(3): $B_R = B_T - B_I$, then (4)

$A_1 (B_T - B_I) - A_1 B_I = -A_2 B_T$

$(A_1 + A_2) B_T = 2 A_1 B_I .$

Q5. continued.

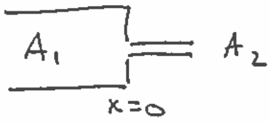
$$B_T = \frac{2A_1}{A_1 + A_2} B_I$$

5

$$\Rightarrow B_R = \frac{2A_1}{A_1+A_2} B_I - B_I = B_I\left(\frac{2A_1 - A_1 - A_2}{A_1+A_2}\right) = B_I\left(\frac{A_1 - A_2}{A_1+A_2}\right)$$

$$|\phi_T|/|\phi_I| = B_T/B_I = \frac{4A_2}{3A_2} = 4/3.$$

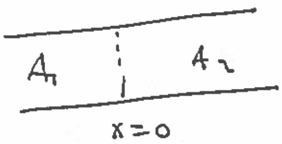
(iv) If $A_2 \ll A_1$



then $B_T \simeq 2B_I$, $B_R = B_I$. incoming wave is reflected with same amplitude.

4

$A_2 \doteq A_1$



then $B_T \simeq B_I$, $B_R \simeq 0$ incoming wave is simply transmitted with little reflection.

20

Two hours

THE UNIVERSITY OF MANCHESTER

MATHEMATICAL BIOLOGY: SOLUTIONS

? May 2019

?.00 – ?.00

Answer **ALL** three questions in Section A (30 marks in total).
Answer **TWO** of the THREE questions in Section B (50 marks in total).
If more than TWO questions from Section B are attempted,
then credit will be given for the best TWO answers.

Electronic calculators may be used in accordance with the University regulations

---

SECTION A

Answer **ALL** three questions

A1. Solution: [this question addresses ILO1 - medium]

(a) r is the net birth minus death rate in the absence of harvesting. K is the carrying capacity of the population in the absence of harvesting. [2 marks]

(b) The ODE reduces to $\mathrm{d}x/\mathrm{d}t = x - x^2 - \alpha \equiv f(x)$. [3 marks]

(c) $f(x)$ is a parabola with maximum $1/4 - \alpha$ at $x = \alpha$. For $0 < \alpha < 1/4$, f has two positive zeros, representing two equilbria. The larger is stable ($f' < 0$) the smaller unstable ($f' > 0$). A sketch can be used to illustrate this. For $\alpha > 1/4$, $\mathrm{d}x/\mathrm{d}t < 0$ for all $x \geq 0$, implying that harvesting eliminates the population. [5 marks]

A2. Solution: [this question addresses ILO3 - medium]

(a) Setting $X = X_0 + z$ and linearizing, noting that

$$(1 + X_0 + z)^{-1} = \frac{1}{1+X_0}\left(1 + \frac{z}{1+X_0}\right)^{-1} \approx \frac{1}{1+X_0}\left(1 - \frac{z}{1+X_0} + \dots\right) \qquad (1)$$

gives

$$\frac{\mathrm{d}z}{\mathrm{d}t} = -\frac{z(t-T)}{(1+X_0)^2}.$$

[3 marks]

(b) Setting $z = ae^{\lambda t}$ requires $(1 + X_0)^2\lambda = -e^{-\lambda T}$. [2 marks]

(c) Setting $\lambda = ib$ for some b requires $\cos bT = 0$ and $(1 + X_0^2)b = \sin bT$. The solution with the smallest T has $bT = \pi/2$ and $b = 1/(1 + X_0)^2$. [3 marks]

(d) The period of oscillation is $2\pi/b = 4T$. [2 marks]

A3. Solution: [this question addresses ILO5 - medium]

(a) Substituting $u = Ae^{\lambda x}$ into $0 = cu_z + Du_{zz} + ru$ gives $0 = f(\lambda)$ where $f(\lambda) = D\lambda^2 + c\lambda + r$. This quadratic has a minimum at $\lambda = -c/(2D)$ and a minimum value $r - (c^2/4D)$. For real λ we need $c^2 \geq 4Dr$, i.e. $c \geq 2\sqrt{Dr}$. [4 marks]

(b) $c = 2\sqrt{Dr}$ implies $\lambda = -c/2D = -\sqrt{r/D}$. This is the lengthscale over which u decays to zero. [3 marks]

(c) Writing $u = 1 + \hat{u}$ and linearising gives $0 = c\hat{u}_z + D\hat{u}_{zz} - r\hat{u}$ and hence $0 = D\mu^2 + c\mu - r = r[\alpha^2 + 2\alpha - 1]$. Thus $\alpha = -1 \pm \sqrt{2}$. To ensure $\hat{u} \to 0$ as $z \to -\infty$, take $\alpha = \sqrt{2} - 1 > 0$. [3 marks]

SECTION B

Answer **TWO** of the three questions

B4. This question addresses ILO1 and ILO5.

(a) $(\mathrm{d}/\mathrm{d}t)(S+I+R)=0$, hence $S+I+R$ is a conserved quantity. We take $S+I+R=N$. As a result $R(t)=N-S(t)-I(t)$ and the R equation need no longer be considered. [2 marks]

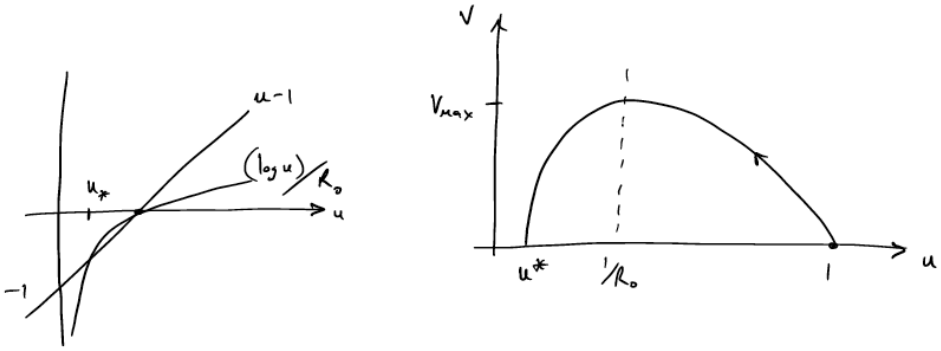
(b) The given equations follow directly with $R_0=\beta N/\gamma$. [5 marks]

(c) Now for $v\neq 0$,

$$\frac{\mathrm{d}v}{\mathrm{d}u}=\frac{v(R_0u-1)}{-R_0uv}=\frac{R_0u-1}{-R_0u}=-1+\frac{1}{R_0u}$$

and hence $v=-u+(1/R_0)\log u+v_0$ for some v_0. Imposing initial conditions, with $v\to 0$ as $u\to 1$, gives $v_0=1$.

$u-1$ and $(\log u)/R_0$ intersect at $u=1$, with slope 1 and $1/R_0$ respectively. Since $R_0>1$, this implies a second intersection in $(0,1)$ — see sketch. [7 marks]



(d) We see that $\mathrm{d}v/\mathrm{d}u=0$ for $u=1/R_0$, passing from positive to negative as u increases through $1/R_0$, hence this is a local maximum. See sketch. The outbreak causes u to fall, v to rise to $v_{\max}$ and then fall, ending with all the infections dying out. $N(1-u^{\star})$ is the number of individuals to have recovered from the disease at the end of the outbreak. [6 marks]

(e)
$$v_{\max} = 1 - \frac{1}{R_0} - \frac{\log R_0}{R_0}$$
so $\mathrm{d}v_0/\mathrm{d}R_0 = \log(R_0)/R_0^2 > 0$ for $R_0 > 1$, so $v_{\max}$ is an increasing function of R_0. [2 marks]

(f) The model is modified as
$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta IS + \nu R, \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta IS - \gamma I, \quad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \nu R.$$
[3 marks]

B5. Solution [this question addresses ILO6] The specific application is unfamiliar - hard.

(a) The governing equations are
$$\begin{aligned}
\dot{S} &= k_{-1}C_1 - k_1SE \\
\dot{E} &= k_{-1}C_1 - k_1SE + k_2C_1 + k_{-3}C_2 - k_3DE \\
\dot{C}_1 &= k_1SE - k_2C_1 - k_{-1}C_1 \\
\dot{C}_2 &= k_3DE - k_{-3}C_2 \\
\dot{D} &= k_{-3}C_2 - k_3DE \\
\dot{P} &= k_2C_1
\end{aligned}$$
[6 marks]

(b) The 6×6 stoichiometric matrix mapping $(k_1SE, k_{-1}C_1, k_2C_1, k_3C_2, k_{-3}C_2)$ to the RHS of the ODEs above is
$$\begin{vmatrix}
-1 & 1 & & & \\
-1 & 1 & 1 & -1 & 1 \\
1 & -1 & -1 & & \\
 & & & 1 & -1 \\
 & & & -1 & 1 \\
 & & 1 & &
\end{vmatrix}\begin{vmatrix}
1 & & & & & \\
 & 1 & & & & \\
 & & 1 & & & \\
 & & & 1 & & \\
 & & & & 1 & \\
 & & & & & 1
\end{vmatrix}$$
Row reduction yields, first,
$$\begin{vmatrix}
-1 & 1 & & & \\
 & & 1 & -1 & 1 \\
 & & & -1 & 1 \\
 & & & 1 & -1 \\
 & & & & \\
 & & 1 & &
\end{vmatrix}\begin{vmatrix}
1 & & & & & \\
-1 & 1 & & & & \\
 & 1 & 1 & & & \\
 & & & 1 & & \\
 & & & 1 & 1 & \\
 & & & & & 1
\end{vmatrix}$$
and then
$$\begin{vmatrix}
-1 & 1 & & & \\
 & & 1 & -1 & 1 \\
 & & & -1 & 1 \\
 & & & & \\
 & & & & \\
 & & & &
\end{vmatrix}\begin{vmatrix}
1 & & & & & \\
-1 & 1 & & & & \\
 & 1 & 1 & & & \\
 & 1 & 1 & 1 & & \\
 & & & 1 & 1 & \\
1 & & 1 & & & 1
\end{vmatrix}$$

The stoichiometric matrix has rank 3, and three conserved quantities are $E + C_1 + C_2 = E_0$, $C_2 + D = D_0$, $S + C_1 + P = S_0$. The reduced equations are therefore

$$\begin{aligned}\dot{C}_1 &= k_1(S_0 - C_1 - P)(E_0 - C_1 - C_2) - k_2C_1 - k_{-1}C_1\\ \dot{C}_2 &= k_3(D_0 - C_2)(E_0 - C_1 - C_2) - k_{-3}C_2\\ \dot{P} &= k_2C_1\end{aligned}$$

[8 marks]

(c) Assuming the complexes equilibrate rapidly gives

$$\begin{aligned}0 &= k_1(S_0 - C_1 - P)(E_0 - C_1 - C_2) - k_2C_1 - k_{-1}C_1\\ 0 &= k_3(D_0 - C_2)(E_0 - C_1 - C_2) - k_{-3}C_2\end{aligned}$$

With abundant S and D we can write

$$\begin{aligned}0 &= k_1S_0(E_0 - C_1 - C_2) - k_2C_1 - k_{-1}C_1\\ 0 &= k_3D_0(E_0 - C_1 - C_2) - k_{-3}C_2\end{aligned}$$

Refomulating in terms of the given parameters gives

$$\begin{aligned}S_0(E_0 - C_1 - C_2) &= LC_1\\ D_0(E_0 - C_1 - C_2) &= KC_2\end{aligned}$$

Thus $C_2 = LD_0C_1/(KS_0)$. Eliminating C_2 gives $KS_0E_0 = C_1[S_0K + D_0L + KL]$ and hence

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{k_2S_0E_0K}{D_0L + S_0K + LK}$$

[8 marks]

(d) $D = 0$ recovers the Michaelis–Menten approximation; $D_0 > 0$ slows this reaction. [3 marks]

B6. This question addresses ILO8.

(a) With $X = X_\star$ and $Y = Y_\star$, each term in both PDEs vanishes and so (1) is satisfied. [2 marks]

(b) Consider the various terms in the PDE for X:

(i) $\partial_t X = \partial_t(X_\star + x(\theta, t)) = \partial_t x$;

(ii) $f(X, Y) = f(X_\star + x, Y_\star + y)$. Expanding this in a Taylor series

$$\begin{aligned}f(X_\star + x, Y_\star + y) &= f(X_\star, Y_\star) + xf_x(X_\star, Y_\star) + yf_y(X_\star, Y_\star) + \dots\\ &\approx 0 + xf_x(X_\star, Y_\star) + yf_y(X_\star, Y_\star)\end{aligned}$$

(iii) $\partial_{\theta\theta}X = \partial_{\theta\theta}(X_\star + x(\theta, t)) = \partial_{\theta\theta}x$

This recovers (2) with

$$a = f_x(X_\star, Y_\star),\quad b = f_y(X_\star, Y_\star),\quad c = g_x(X_\star, Y_\star)\ \text{ and }\ d = g_y(X_\star, Y_\star).$$

[8 marks]

(c) If $x(\theta, t) = u(t)\cos(n\theta)$ then

$$\partial_t x = \dot{u}\cos(n\theta) \qquad \text{and} \qquad \partial_{\theta\theta} x = -n^2 u \cos(n\theta).$$

Putting these into the PDE and dividing by $\cos(n\theta)$ produces the desired form. [4 marks]

(d) The ODEs for $\dot{u}$ and $\dot{v}$ are linear, so their asymptotic behaviour depends on the eigenvalues of the matrix

$$A = \begin{bmatrix} (a - \mu n^2) & b \\ c & (d - \nu n^2) \end{bmatrix}.$$

If any of them has a positive real part, then the spatially uniform solution is unstable to oscillatory patterns with spatial wavelength $\lambda = 2\pi/n$. The eigenvalues of A satisfy

$$\lambda^2 - \Sigma\lambda + \Delta = 0$$

where

$$\Sigma = \mathrm{Tr}(A) = a + d - n^2(\mu + \nu) \qquad \text{and} \qquad \Delta = \det A = (a - \mu n^2)(d - \nu n^2) - bc.$$

For stability we need $\Sigma < 0$ and $\Delta > 0$. Thus when $n = 0$, the conditions ensuring stability are

$$a + d < 0 \quad \text{and} \quad ad - bc > 0. \tag{2}$$

[6 marks]

(e) For $n > 0$, the condition ensuring stability is

$$a + d - n^2(\mu + \nu) < 0 \qquad \text{or} \qquad (a - \mu n^2)(d - \nu n^2) - bc > 0 \tag{3}$$

for all values of n. The first condition is always satisfied. The second may be written

$$ad - bc > (a\nu + d\mu)n^2 - \mu\nu n^4.$$

Let $f(z) = (a\nu + d\mu)z - \mu\nu z^2$. Then $f'(z) = (a\nu + d\mu) - 2\mu\nu z$, so that the maximum is attained at $z = (a\nu + d\mu)/(2\mu\nu)$, and

$$f_{\max} = \frac{(a\nu + d\mu)^2}{4\mu\nu}. \tag{4}$$

Thus stability is guaranteed if $ad - bc > f_{\max}$, i.e.

$$ad - bc > \frac{\mu\nu}{4}\left(\frac{a}{\mu} + \frac{d}{\nu}\right)^2$$

as required.

[5 marks]

END OF EXAMINATION PAPER

TLOs

1 apply the orbit-stabilizer theorem,
2 determine the Burnside type of a finite group action,
3 perform calculations using the Seitz symbol for a Euclidean transformation,
4 recognize the 5 types of symmetry of a plane lattice, and use this to analyze wallpaper groups,
5 find the symmetric solutions to problems with symmetry,
6 predict the possible symmetries of periodic motions in symmetric systems.

Question 1

a) i)  $\to D_2$ order 4 Abelian

ii)  $\to D_4$ order 8 not Abelian

iii)  D_3 6 "

ILO 1 (?) elementary

---

b) $f(x)$: is clearly symmetric about the y-axis & periodic with period 2:
Symmetry group generated by $T_1 x = -x$
& $T_2 x = x+2$.

ILO 1 easy

$g(x)$: Again periodic with period 2,
also $g(x+1) = -g(x)$. So full symmetry group generated by $Tx = x+1$.

Question 2 $\quad G$ acts on X. 1201 mixed levels

i) Orbit of x is $G\cdot x = \{ g\cdot x \mid g \in G\}$

Stabilizer is $G_x = \{ g \in G \mid g\cdot x = x\}$.

Let $g, h \in G_x$. Then
$$(gh)\cdot x = g\cdot(h\cdot x) = g\cdot x = x$$
and $h^{-1}\cdot x = h^{-1}(h\cdot x) = (h^{-1}h)\cdot x = e\cdot x = x$

Hence $gh \in G_x$ and $h^{-1} \in G_x$.

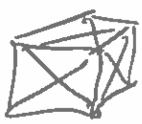
Thus G_x satisfies the subgroup criterion. $\boxed{4}$

ii) The orbit-stabilizer theorem states $\boxed{2}$
$$|G\cdot x| = |G|/|G_x|.$$

iii) Let $y \in G\cdot x$. Then
$$\begin{aligned} G_{x,y} &= \{ g \in G \mid g\cdot x = y\} \\ &= \{ g \in G \mid g\cdot x = h\cdot x\} \\ &= \{ g \in G \mid h^{-1}g \in G_x\} \\ &= hG_x. \end{aligned}$$
bookwork

The orbit-stabilizer theorem then follows by showing the map
$$\begin{aligned} G\cdot x &\to G/G_x \\ y &\mapsto G_{x,y} \end{aligned}$$
is a bijection. $\boxed{4}$

unseen

(iv)  The set of diagonals is a group orbit and there are $6\times 2 = 12$ elements. So $|G_x| = |G|/|G\cdot x| = 24/12 = 2$. $\boxed{2}$

2 (b) Sufficient to show equivariance for the generators:

[ILO 5]

Unseen example

$r_0(x,y) = (x,-y)$, so

$$f \circ r_0 (x,y) = (x - x(-y)^2, (-y) - x^2(-y))$$
$$= (x - xy^2, -y + x^2 y)$$
$$= (x - xy^2, -(y - x^2 y))$$
$$= r_0 \circ f(x,y).$$

Similarly,

$R_{\pi/2}(x,y) = (-y,x)$, and

$$f \circ R_{\pi/2}(x,y) = f(-y,x) = (-y + yx^2, x - y^2 x)$$
$$= R_{\pi/2} \circ f(x,y)$$

---

[4]

Question 3

7LO 384

a) $(A|\underline{v})\underline{x} = A\underline{x} + \underline{v}$ – definition of Seitz symbol.

Bookwork (asked every year)

Thus

$$(A|\underline{v})(B|\underline{u})\underline{x} = (A|\underline{v})(B\underline{x} + \underline{u})$$

$$= A(B\underline{x} + \underline{u}) + \underline{v}$$

$$= AB\underline{x} + (A\underline{u} + \underline{v})$$

$$= (AB|A\underline{u} + \underline{v}).\underline{x}.$$

This holds $\forall \underline{x} \in \mathbb{R}^2$, whence

$$(A|\underline{v})(B|\underline{u}) = (AB|\underline{v} + A\underline{u}).$$ $\boxed{4}$

Let $(B|\underline{u}) = (A|\underline{v})^{-1}$. We require

Bookwork

$$(AB|\underline{v} + A\underline{u}) = (I|\underline{0})$$

i.e. $AB = I$

$\underline{v} + A\underline{u} = 0$

$\therefore \quad B = A^{-1}, \quad \underline{u} = -A^{-1}\underline{v}$

Thus, $(A|\underline{v})^{-1} = (A^{-1}|-A^{-1}\underline{v})$. $\boxed{2}$

b) $L = \{(x,y) \in \mathbb{R}^2 \mid x \in \mathbb{Z}, y \in \mathbb{Z}, x+y \in 3\mathbb{Z}\}$.

i) $\underline{u}_1 = (3,0) \in L$ because

$0, 3 \in \mathbb{Z}$ & $x+y = 3 \in 3\mathbb{Z}$ ✓

seen similar

$\underline{u}_2 = (0,3) \in L$ because

$0, 3 \in \mathbb{Z}$ & $x+y = 3 \in 3\mathbb{Z}$ ✓

$(1,2) \in L$

since $x+y = 3 \in 3\mathbb{Z}$

but

$(1,2) \notin \mathbb{Z}\{\underline{u}_1, \underline{u}_2\}$

(ii) Try $\underline{a} = (1,-1)$, $\underline{b} = (2,1)$.

seen similar but takes some trial and error.

First show $L = \mathbb{Z}\{\underline{a}, \underline{b}\}$.

Clearly $\underline{a}, \underline{b} \in L$ since $\pm 1, 2 \in \mathbb{Z}$ and

(for $\underline{a}$) $x+y = 0$
(for $\underline{b}$) $x+y = 3$ $\Big\} \in 3\mathbb{Z}$ ✓

$\therefore \mathbb{Z}\{\underline{a}, \underline{b}\} \subset L$ (given that L is a lattice).

Conversely, given $(x,y) \in L$ we want to find integers $m, n \in \mathbb{Z}$ such that

$$(x,y) = m\underline{a} + n\underline{b}.$$

That is $\quad x = m + 2n$
$\quad\quad\quad\quad y = -m + n$

($m, n \in \mathbb{R}$ exist + are unique since $\{\underline{a}, \underline{b}\}$ is a basis)

Then

$$\begin{cases} x+y = 3n \\ 2x - y = 3m \end{cases}$$

$$\therefore \quad n = \frac{x+y}{3} \in \mathbb{Z} \checkmark$$

$$m = \frac{2x-y}{3} = x - y - \frac{x+y}{3} \in \mathbb{Z} \checkmark$$

Therefore $\quad L \subset \mathbb{Z}\{\underline{a}, \underline{b}\}$.

Now show $\underline{a}$ is shortest
$\quad\quad\quad\quad \underline{b}$ next shortest $\quad$ etc

Listing points in $[-3,3] \times [-3,3]$ in diagram in terms of $\underline{a}, \underline{b}$:

	$\underline{0}$,	$\underline{a}$,	$-\underline{a}$,	$\underline{b} = (2,1)$,	$\underline{b} - \underline{a} = (1,2)$	$\underline{a} + \underline{b} = (3,0)$
Lengths:	0	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{5}$	3

$$\therefore \quad |\underline{a}| < |\underline{b}| = |\underline{b} - \underline{a}| < |\underline{b} + \underline{a}|$$

and the lattice is centred Rectangular (= rhombic).

Question 4

$$\begin{cases} \dot{x} = x - yz \\ \dot{y} = y - xz \\ \dot{z} = z - xy \end{cases}$$

1605

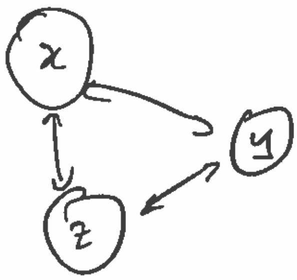
level: medium

a) S_3 is generated by transposition $(1\,2)$ & $(1\,3)$ say. Applying $(1\,2)$ means swapping x & y: system becomes

$$\begin{cases} \dot{y} = y - xz \\ \dot{x} = x - yz \\ \dot{z} = z - yx \end{cases}$$

easy

which is identical. Swapping x & z is similar.



Cell diagram.

4

b) The diagonal is $\mathrm{Fix}(S_3, \mathbb{R}^3)$ and as such it's invariant under the dynamics.

Thm: If a system of o.d.e.'s on $\mathbb{R}^n$ has symmetry G & $H < G$, then $\mathrm{Fix}(H, \mathbb{R}^n)$ is invariant under the dynamics.

4

c) Equilibrium points:

$$\begin{cases} x - yz = 0 \\ y - xz = 0 \\ z - xy = 0 \end{cases}$$

solve first for $x = yz$ & substitute into others two:
$$\begin{cases} y - yz^2 = 0 \\ z - y^2 z = 0 \end{cases}$$

ie
$$\begin{cases} y(1-z^2) = 0 \\ z(1-y^2) = 0 \end{cases}$$

Solutions:

$y = z = 0 \Rightarrow x = 0$ $\qquad (0, 0, 0)$

$y = 1,\ z = 1 \Rightarrow x = 1$ $\qquad (1, 1, 1)$

$y = 1,\ z = -1 \Rightarrow x = -1$ $\qquad (-1, 1, -1)$

$y = -1,\ z = 1 \Rightarrow x = -1$ $\qquad (-1, -1, 1)$

$y = -1,\ z = -1 \Rightarrow x = 1$ $\qquad (1, -1, -1)$.

5 equilibria.

The orbits are

$\{(0,0,0)\}$, $\{(1,1,1)\}$, $\{(1,-1,-1), (-1,1,-1), (-1,-1,1)\}$

The first two are fixed by S_3

The others are fixed by (a conjugate of) $H = \langle (1\,2) \rangle$

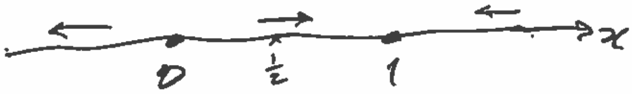
Burnside type is

$$2(S_3) + 1(H).$$

medium

6

d) The diagonal is a line & can be parametrized by x, say Unseen



On that line $y=z=x$, so Harder

$$\dot{x} = x - x^2$$

If $x \in (0,1)$ then $\dot{x} > 0$, so trajectory starting at $x = 1/2$ must increase and limit on $x = 1$.

ie $\lim_{t\to\infty} \underline{x}(t) = (1,1,1)$

$\boxed{4}$

<u>Question 5</u>: 1206 medium

a) $S^1 = \mathbb{R}/\mathbb{Z}$

$$\Sigma_\gamma = \{ (g, \theta) \in G \times S^1 \mid \gamma(t + \theta T) = g \cdot \gamma(t) \}$$

3

b) $\tilde{C}_2 = \{ (I, 0),\ (R_\pi, \tfrac{1}{2}) \} \subset G \times S^1$

$\tilde{D}_1 = \{ (I, 0),\ (r_0, \tfrac{1}{2}) \}$

bookwork

Let $\gamma(t) = (x(t), y(t))$.

Then $\Sigma_\gamma = \tilde{C}_2$ means

$$\gamma(t + T/2) = -\gamma(t)$$

$$\therefore \begin{cases} x(t + T/2) = -x(t) \\ y(t + T/2) = -y(t) \end{cases}$$

And $\Sigma_\gamma = \tilde{D}_1$ means

$$\begin{cases} x(t + T/2) = x(t) & \leftarrow \text{period } T/2 \\ y(t + T/2) = -y(t). \end{cases}$$

unseen part

<u>Experiment 1</u>: dashed line has period $T/2$
solid line is symmetric
about y-axis after translation
by $T/2$

$\longrightarrow \tilde{D}_1$.

<u>Experiment 2</u>: both satisfy $u(t + T/2) = -u(t)$
so $\Sigma = \tilde{C}_2$.

7

The **intended learning outcomes (ILOs)** of the course are as follows:

ILO1 Characterise the best approximation of a function using different norms.

ILO2 Compute Padé approximations and evaluate their quality.

ILO3 Derive quadrature rules and their error bounds.

ILO4 Apply the Trapezium rule, Gauss quadrature and adaptive quadrature to compute integrals.

ILO5 Describe the Romberg scheme in the context of extrapolation.

ILO6 Analyse and apply one-step, multi-step, and the Euler method for solving ordinary differential equations (ODE).

ILO7 Solve ODEs numerically using Runge-Kutta, Trapezium and higher-order methods.

ILO8 Quantify the error and convergence of numerical solvers for ODEs.

ILO9 Recognize some of the difficulties that can occur in the numerical solution of problems arising in science and engineering.

All questions on the exam paper address one or more of these ILOs. The associated level of difficulty (at which the ILO is being assessed) is indicated with **H** (high), **M** (medium) or **L** (low).

SECTION A

The material in this section is considered 'core' and is all bookwork (from lecture notes) or else very similar to questions posed on the Examples Sheets (for which full worked solutions were given).

A1. This question tests **ILO1, L(a–c)**.

(a) We have $\| f \|_{2,w} := \left(\int_a^b w(x) f(x)^2 dx \right)^{1/2}$ where $w(x)$ is non-negative and continuous on (a, b) with $\int_a^b w(x)dx > 0$. **[2 marks]** We say g is a best $L_{2,w}$ approximation to f from the set G if

$$\| f - g \|_{2,w} \leq \| f - h \|_{2,w}, \qquad \forall h \in G.$$

[2 marks]

(b) The best $L_{2,w}$ approximation of this form is found by solving the normal equations $A\mathbf{c} = \mathbf{f}$ (for the coefficients $\mathbf{c} = [c_0, \ldots, c_n]^\top$) where

$$A_{i,j} = \langle \phi_i, \phi_j \rangle_w, \quad f_j = \langle f, \phi_j \rangle_w, \qquad i, j = 0, 1, \ldots, n,$$

where $< u, v >_w = \int_a^b w(x)u(x)v(x)\, dx$. **[3 marks]** When the polynomials ϕ_i are orthogonal, we have $\langle \phi_i, \phi_j \rangle_w = 0$ for $i \neq j$ so A is diagonal and the system is trivial to solve. We have,

$$c_i = \frac{\langle f, \phi_i \rangle_w}{\langle \phi_i, \phi_i \rangle_w}, \qquad i = 1, 2, \ldots, n.$$

[2 marks]

A2. This question tests **ILO1, L(a), M(b)**.

The cosine function has roots at $x = (2i+1)\pi/2, i \in \mathbb{Z}$, so the $n+1$ zeros of $T_{n+1}(x)$ on $[-1,1]$, (the so-called *Chebyshev points*), are

$$x_i = \cos\left(\frac{(2i+1)\pi}{2(n+1)}\right), \qquad i = 0, 1, \ldots, n.$$

[**3 marks**] It is known (due to the Chebyshev Equioscillation Theorem) that among all monic polynomials of degree $n+1$, $2^{-n}T_{n+1}(x)$ has the smallest L_∞-norm on $[-1,1]$. The error in the polynomial $p_n(x)$ of degree $\le n$ that interpolates to $f(x)$ at a set of $n+1$ distinct points $\{x_i\}_{i=0}^n$ in $[-1,1]$ has the form

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}\prod_{i=0}^{n}(x - x_i).$$

Since the factor $\prod_{i=0}^n(x - x_i)$ is a monic polynomial of degree $n+1$, its L_∞-norm is minimized by choosing the interpolation points x_i to be the zeros of T_{n+1}. [**4 marks**]

A3. This question tests **ILO3(L)** and **ILO5(L)**.

(a) Doubling the number of intervals corresponds to reducing h by a factor of 2 [**1 mark**]. If we replace h by $h/2$ in the asymptotic formula, we get

$$I(f) - T(h/2) = a_2\frac{h^2}{4} + a_4\frac{h^4}{16} + \mathcal{O}(h^6).$$

The new scheme $T(h/2)$ still has an error that is $\mathcal{O}(h^2)$ but the leading term in the expression for the error is reduced by a factor of four. [**2 marks**]

(b) We have

$$4I(f) - 4T(h/2) = a_2h^2 + a_4\frac{h^4}{4} + \mathcal{O}(h^6)$$

and

$$I(f) - T(h) = a_2h^2 + a_4h^4 + \mathcal{O}(h^6).$$

Hence,

$$I(f) - \left(\frac{4T(h/2) - T(h)}{3}\right) = -\frac{a_4h^4}{4} + \mathcal{O}(h^6).$$

The approximation $\frac{4}{3}T(h/2) - \frac{1}{3}T(h)$ gives an error that is $\mathcal{O}(h^4)$. [**4 marks**].

A4. This question tests **ILO9(M)** (and **ILO3(M)**).

Let $f_i = f(x_i)$ and set $\bar{f}_i = f_i + \epsilon_i$, with $|\epsilon_i| \le 10^{-4}$. Writing the first rule (Simpson's rule) as $J(f) = \frac{1}{6}(f_0 + 4f_{\frac{1}{2}} + f_1)$ we have

$$|J(f) - J(\bar{f})| = \frac{1}{6}|\epsilon_0 + 4\epsilon_{\frac{1}{2}} + \epsilon_1| \le \frac{1}{6}(10^{-4} + 4\cdot 10^{-4} + 10^{-4}) = 10^{-4}.$$

[**2 marks**]. Similarly, for the open rule,

$$|J(f) - J(\bar{f})| \le \frac{1}{3}(2\cdot 10^{-4} + 10^{-4} + 2\cdot 10^{-4}) = \frac{5}{3}\cdot 10^{-4}.$$

[**2 marks**]. So, for the closed rule, where all the weights are positive, errors of size at most 10^{-4} in the f_i values change the approximation to the integral by at most 10^{-4}. But for a rule with weights of both sign, the change in the quadrature rule estimate can exceed 10^{-4} (the upper bound for the data error). [**3 marks**].

A5. This question tests **ILO6, L(a–b)**.

(a) To derive the ℓ-step A–B method, we replace the integrand in

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x))\, dx,$$

by the polynomial $p(x)$ of degree $\ell - 1$ which interpolates to the values $f_n, f_{n-1}, \ldots, f_{n+1-\ell}$ at the ℓ points $x_n, x_{n-1}, \ldots, x_{n+1-\ell}$. Similarly, to derive the ℓ-step A–M method, we replace the integrand by the polynomial $p(x)$ of degree ℓ which interpolates to the values $f_{n+1}, f_n, f_{n-1}, \ldots, f_{n+1-\ell}$ at the $\ell + 1$ values $x_{n+1}, x_n, x_{n-1}, \ldots, x_{n+1-\ell}$. [**2 marks**]
The Adams-Bashforth method is explicit (has $b_0 = 0$) and has order ℓ. The Adams-Moulton method is implicit but has order $\ell + 1$. Hence the A–B methods are easier to implement but less accurate; the A–M methods are harder to implement (as a nonlinear equation needs to be solved for y_{n+1}) but have better accuracy. [**4 marks**]

(b) To derive the 1-step AB method, we replace the integrand with the polynomial $p_0(x)$ of degree zero which takes the value f_n at x_n. Setting $p_0 = f_n$, the resulting approximation scheme is:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f_n\, dx = y_n + h f_n.$$

This is the standard (forward) Euler method. [**4 marks**]

SECTION B

B6. Parts (a) and (c) are bookwork (from the lecture notes) and part (b) was posed as a question on an examples sheet. This question tests **ILO2, L(a), L+M(b), H(c)**.

(a) Polynomials cannot have asymptotes and they are always finite on the finite real axis and tend to $\pm\infty$ as $x \to \pm\infty$. They also have a tendency to oscillate. A Padé approximant is a ratio of polynomials. A rational function with equal degree numerator and denominator stays bounded as $x \to \pm\infty$. A rational function also has poles (the roots of the denominator polynomial). Rationals can also be free of oscillations. [**2 marks for any two**]. We could evaluate a Padé approximant using Horner's method, applied to the numerator and denominator separately. However, the continued fraction representation is often cheaper to evaluate. More importantly, evaluating the continued fraction tends to be more stable (and leads to smaller errors in the evaluation). [**2 marks**]

(b) The Taylor series expansion of $f(x) = \log(1+x)$ about $x = 0$ is

$$f(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

[**2 marks**] and the first few coefficents (given) in $r_n(x)$ are $c_1 = 1, c_2 = 1/2, c_3 = 1/6, c_4 = 1/3, \dots$. Hence

$$r_3(x) = \frac{x}{1 + \frac{x/2}{1+x/6}} = \frac{x + x^2/6}{1 + 2x/3}.$$

[**2 marks**] We see that $r_3(x)$ is the ratio of a quadratic polynomial $p_{21}(x) = x + x^2/6$ and the linear polynomial $q_{21} = 1 + 2x/3$ with $q_{21}(0) = 1$. [**1 mark**] We also have

$$\begin{aligned} f(x)q_{21}(x) - p_{21}(x) &= \left(x - \frac{x^2}{2} + \frac{x^3}{3} + O(x^4)\right)(1 + 2x/3) - (x + x^2/6) \\ &= x + 2x^2/3 - x^2/2 - 2x^3/6 + x^3/3 + O(x^4) - x - x^2/6 = O(x^4). \end{aligned}$$

[**2 marks**]. Hence $r_3(x)$ is the [2/1] Padé approximant, as claimed.

(c) We show by induction that $r_n(x) = p_n/q_n$, where p_n, q_n are computed by the stated algorithm. First, let $n = 1$ (base case). We have

$$\frac{p_1}{q_1} = \frac{b_1 b_0 + a_1 x p_{-1}}{b_1 q_0 + a_1 x q_{-1}} = \frac{b_1 b_0 + a_1 x}{b_1} = b_0 + \frac{a_1 x}{b_1} =: r_1(x).$$

Hence, the result is true for $n = 1$. [**3 marks**] Assume the result is true for $n = k$, that is,

$$r_k(x) = \frac{p_k}{q_k} = \frac{b_k p_{k-1} + a_k x p_{k-2}}{b_k q_{k-1} + a_k x q_{k-2}}. \tag{1}$$

To obtain $r_{k+1}(x)$ from $r_k(x)$ we simply replace b_k by $b_k + a_{k+1}x/b_{k+1}$. [**1 marks**] Making this substitution in the right-hand side of (1) gives

$$\begin{aligned} \frac{(b_k + a_{k+1}x/b_{k+1})p_{k-1} + a_k x p_{k-2}}{(b_k + a_{k+1}x/b_{k+1})q_{k-1} + a_k x q_{k-2}} &= \frac{(b_k b_{k+1} + a_{k+1}x)p_{k-1} + a_k x b_{k+1} p_{k-2}}{(b_k b_{k+1} + a_{k+1}x)q_{k-1} + a_k x b_{k+1} q_{k-2}} \\ &= \frac{b_{k+1}(b_k p_{k-1} + a_k x p_{k-2}) + a_{k+1} x p_{k-1}}{b_{k+1}(b_k q_{k-1} + a_k x q_{k-2}) + a_{k+1} x q_{k-1}} \\ &= \frac{b_{k+1} p_k + a_{k+1} x p_{k-1}}{b_{k+1} q_k + a_{k+1} x q_{k-1}} = \frac{p_{k+1}}{q_{k+1}}, \end{aligned}$$

which is the result with k replaced by $k+1$, which completes the proof. [**5 marks**]

B7. Parts (a) and (b) are bookwork (from the lecture notes) and part (c) was posed as a question on an examples sheet. This question tests **ILO3** and **ILO4, L(a), M(b,c)**.

(a) If $p_{n-1}(x)$ is the polynomial of degree $n-1$ or less that interpolates f at the n chosen points x_i, we have (with $f_i = p_{n-1}(x_i)$)

$$\int_a^b w(x)p_{n-1}(x)\,dx = \int_a^b w(x)\sum_{i=1}^n f_i l_i(x)\,dx = \sum_{i=1}^n f_i \int_a^b w(x)l_i(x)\,dx,$$

where $l_i(x)$ is the Lagrange interpolating polynomial of degree $n-1$ satisfying $l_i(x_j) = \delta_{ij}$. Hence, to make the rule exact for polynomials of degree up to $n-1$, we should take

$$w_i = \int_a^b w(x)l_i(x)\,dx.$$

[4 marks]

(b) Let f be a polynomial of degree $\leq 2n-1$. Write $f(x) = q(x)\phi_n(x) + r(x)$, where q and r are polynomials of degrees $\leq n-1$. Then

$$\begin{aligned} I(f) &= \underbrace{\int_a^b w(x)q(x)\phi_n(x)\,dx}_{\substack{=0\text{ by orthogonality}\\ \text{since } q(x)=\sum_{i=0}^{n-1}\alpha_i\phi_i}} + \int_a^b w(x)r(x)\,dx, \\ G_n(f) &= \underbrace{\sum_{i=1}^n w_i q(x_i)\phi_n(x_i)}_{\substack{=0\text{ because the } x_i\\ \text{are the zeros of } \phi_n}} + \sum_{i=1}^n w_i r(x_i). \end{aligned}$$

Now $\int_a^b w(x)r(x)\,dx = \sum_{i=1}^n w_i r(x_i)$, by the choice of the weights w_i, and since r has degree $\leq n-1$, so $I(f) = G_n(f)$, as required. **[8 marks]**

(c) Setting $f(x) = 1, x, x^2, x^3$ gives

$$\begin{aligned} f(x)\equiv 1 &\;\Rightarrow\; \sqrt{\pi} = w_1 + w_2, &(2)\\ f(x)\equiv x &\;\Rightarrow\; 0 = w_1x_1 + w_2x_2, &(3)\\ f(x)\equiv x^2 &\;\Rightarrow\; \tfrac{1}{2}\sqrt{\pi} = w_1x_1^2 + w_2x_2^2, &(4)\\ f(x)\equiv x^3 &\;\Rightarrow\; 0 = w_1x_1^3 + w_2x_2^3, &(5) \end{aligned}$$

where in (i) and (iii) we have used the given results and in (ii) and (iv) the integrand is an odd function. **[4 marks]**. Let $\Phi(x) = (x-x_1)(x-x_2) = x^2 + ax + b$ (following the given HINT). Then

$$\begin{aligned} b\times(2) + a\times(3) + (4) &\;\Rightarrow\; b\sqrt{\pi} + \tfrac{1}{2}\sqrt{\pi} = 0 \quad\Rightarrow\quad b = -\tfrac{1}{2},\\ b\times(3) + a\times(4) + (5) &\;\Rightarrow\; \tfrac{1}{2}a\sqrt{\pi} = 0 \quad\Rightarrow\quad a = 0. \end{aligned}$$

Therefore x_1, x_2 are the roots of $x^2 - \frac{1}{2} = 0$, that is, $x_1 = -x_2 = \sqrt{1/2}$. Then (3) yields $w_1 = w_2 = \sqrt{\pi}/2$ by (2). Hence the two-point rule is

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)\, dx \approx \frac{\sqrt{\pi}}{2}\left(f\left(-\frac{1}{\sqrt{2}}\right) + f\left(\frac{1}{\sqrt{2}}\right)\right).$$

[4 marks]

B8. Parts (a) and (b) are book work (from the lecture notes). The calculations in (c) and (d) were covered in lectures. A hint is given in (d) and (c) is included to enable all students to attempt (d). This question tests **ILO7, L(b,c), M(a), M(d)**.

(a) Consider the interval $[x_n, x_{n+1}]$ and assume we have computed y_n at the grid point x_n. Take one step of Euler's method to generate an approximation at x_{n+1} but regard this as only a first estimate. That is,

$$\widehat{y}_{n+1} = y_n + hf(x_n, y_n).$$

We can now evaluate the derivative at the *end* of the Euler step, using this trial value, $y'(x_{n+1}) \approx f(x_{n+1}, \widehat{y}_{n+1})$. Instead of accepting the Euler estimate, we complete the step to x_{n+1} by using the *average* of the two slopes at the beginning and end of the step:

$$y_{n+1} = y_n + \frac{h}{2}\Big(f(x_n, y_n) + f(x_{n+1}, \widehat{y}_{n+1})\Big).$$

This is the stated scheme. [**3 marks**]
We can also write this method in the following form

$$k_1 = f(x_n, y_n), \quad k_2 = f(x_n + h, y_n + hk_1), \quad y_{n+1} = y_n + \frac{h}{2}(k_1 + k_2).$$

This is an example of a 2-stage Runge–Kutta method,

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + c_2 h, y_n + h a_{21} k_1), \\ y_{n+1} &= y_n + h(b_1 k_1 + b_2 k_2) \end{aligned}$$

with $c_2 = a_{21} = 1$ and $b_1 = b_2 = 1/2$. [**3 marks**]

(b) The local truncation error $\tau(h)$ is the remainder when the exact solution $y(x_n)$ is substituted for y_n in the approximation scheme. If $\tau(h) = \mathcal{O}(h^{p+1})$ then we say the method has order p. [**3 marks**]

(c) Substituting the exact solution $y(x_n)$ for y_n in equation (1) and subtracting the right-hand side from the left-hand side gives:

$$\tau(h) = y(x_{n+1}) - y(x_n) - \frac{h}{2}\big(f(x_n, y(x_n)) + f(x_n + h, y(x_n) + hf(x_n, y(x_n)))\big).$$

[**2 marks**] Noting that $f(x_n, y(x_n)) = y'(x_n)$ then gives

$$\tau(h) = y(x_{n+1}) - y(x_n) - \frac{h}{2}\big(y'(x_n) + f(x_n + h, y(x_n) + hy'(x_n))\big).$$

[**1 mark**]

(d) To prove the method is order two, we need to show that $\tau(h)$ is $\mathcal{O}(h^3)$ (i.e., that the terms in h^0, h and h^2 cancel out). Using a Taylor series expansion for $y(x_{n+1})$ and the hint gives

$$\begin{aligned} y(x_{n+1}) &= y(x_n) + hy'(x_n) + \frac{h^2}{2}y''(x_n) + \mathcal{O}(h^3) \\ &= y(x_n) + hy'(x_n) + \frac{h^2}{2}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}f\right)\Big|_{(x_n,y(x_n))} + \mathcal{O}(h^3). \end{aligned}$$

[3 marks] Next, using a Taylor series in two dimensions:

$$\begin{aligned} f(x_n+h, y(x_n)+hy'(x_n)) &= f(x_n, y(x_n)) + h\frac{\partial f}{\partial x}(x_n, y(x_n)) + hy'(x_n)\frac{\partial f}{\partial y}(x_n, y(x_n)) + \mathcal{O}(h^2) \\ &= y'(x_n) + h\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}f\right)\Big|_{(x_n,y(x_n))} + \mathcal{O}(h^2) \end{aligned}$$

[3 marks] Substituting both expansions into the expression for $\tau(h)$ in part (c) gives

$$\begin{aligned} \tau(h) &= \left[y(x_n) + hy'(x_n) + \frac{h^2}{2}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}f\right)\Big|_{(x_n,y(x_n))} + \mathcal{O}(h^3)\right] - y(x_n) - \frac{h}{2}y'(x_n) \\ &\quad -\frac{h}{2}\left(y'(x_n) + h\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}f\right)\Big|_{(x_n,y(x_n))} + \mathcal{O}(h^2)\right) = \mathcal{O}(h^3). \end{aligned}$$

Hence, the method has order 2. **[2 marks]**

ILO MATH37012

Upon successful completion, the students are expected to able to:

1. classify the states of a discrete time Markov chain;

2. calculate the stationary and limiting distributions of discrete time Markov chains;

3. state and prove results about recurrence, transience and periodicity of discrete time Markov chains;

4. write down and solve the forward equations for simple birth- death processes;

5. apply continuous time Markov chain theory to the modelling of queues;

6. state and prove results about the limiting behaviour of continuous time Markov chains.

Notes for external examiner.

BW=bookwork PS=problem sheet SS=seen similar US=unseen

For the first time, all questions are compulsory, rather than a choice of 3 from 4.

Q1 ILO1,ILO2,ILO3 low

a) i), ii) BW iii) US iv) US will find difficult v) BW
b) i) SS ii) SS iii) first part SS, second US iv) SS

Q2 ILO2,ILO3 medium

i) SS ii) SS iii) PS iv) US v) BW, SS on PS but will find last part difficult.

Q3 ILO4,ILO6 medium

a) i), ii) BW
b) i) PS ii) US will find difficult.

Q4 ILO5 medium/high, ILO6 low

a) i), ii) PS
b) US will probably find iii) tricky.

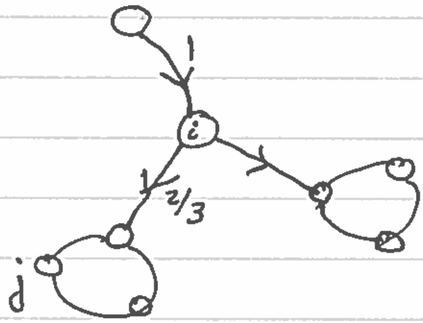
MATH37012 solutions

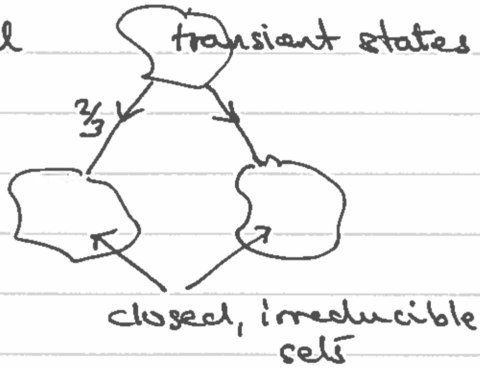
1. a)(i) A state i is <u>recurrent</u> if

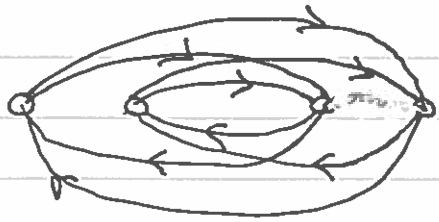
$$P(X_n = i \text{ for some } n \geq 1 \,/\, X_0 = i) = 1.$$

Otherwise, it is <u>transient.</u>

(ii) $\underline{\pi}$ is a stationary distribution for a M.C. with transition matrix P if
(i) $\pi_i \geq 0 \ \forall i \in S$ (ii) $\underline{\pi} P = \underline{\pi}$ (iii) $\sum_{i \in S} \pi_i = 1.$

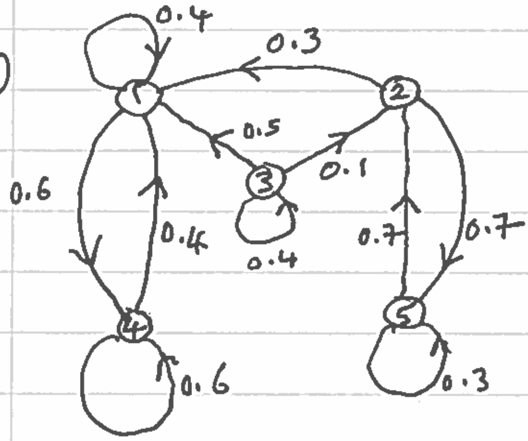
(iii) 

In general 

(iv) Irreducible, period 2 

(v) trivially 

More generally, at least two closed, irreducible sets included.

b) 

$\{X_0 = 5, X_1 = 2, X_2 = 1\} \subseteq \{X_0 = 5, \forall n \geq 1, X_n \neq 5\}$

Therefore

$0 \leq P(X_2 = 1, X_1 = 2 / X_0 = 5) \leq 1 - f_{55}$

$\Rightarrow 0.3 \times 0.7 \leq 1 - f_{55}$

$\Rightarrow f_{55} \leq 0.79 < 1$

and state 5 is therefore transient.

MATH 37012 Solutions

1. b) (cont.) Writing out $\pi P = \pi$ as a system of equations, we have

$$
\begin{aligned}
0.4\pi_1 + 0.3\pi_2 + 0.5\pi_3 + 0.4\pi_4 &= \pi_1 \quad -(1)\\
0.1\pi_3 + 0.7\pi_5 &= \pi_2 \quad -(2)\\
0.4\pi_3 &= \pi_3 \quad -(3)\\
0.6\pi_1 + 0.6\pi_4 &= \pi_4 \quad -(4)\\
0.7\pi_2 + 0.3\pi_5 &= \pi_5 \quad -(5)
\end{aligned}
$$

(3) $\Rightarrow \pi_3 = 0$ then (5) $\Rightarrow \pi_2 = \pi_5$; (2) $\Rightarrow \pi_2 = \pi_5 = 0$;
Then (1) and (4) give

$$0.4\pi_1 + 0.4\pi_4 = \pi_1 \quad -(6)$$
$$0.6\pi_1 + 0.6\pi_4 = \pi_4 \quad -(7)$$

satisfied by $\pi_4 = \frac{3}{2}\pi_1$. Then $\sum \pi_i = 1 \Rightarrow \pi_1 + \pi_4 = 1$
$\Rightarrow \pi_1 + \frac{3}{2}\pi_1 = 1 \Rightarrow \pi_1 = 2/5, \pi_4 = 3/5$.

Summarising A stationary distribution is $(2/5, 0, 0, 3/5, 0)$.

Could also argue that, for transient states, $\pi_i = 0$ and then solve (6) & (7) considering $\{1, 4\}$ as a 2 state M.C

(iii) $\{1, 4\}$ is a closed irreducible set, so by result in notes,

$$\lim_{n\to\infty} p_{11}(n) = \pi_1 = 2/5 \quad \text{as found from (6), (7)}$$

Can quote: If j is transient, then $\forall i \in S$ $p_{ij}(n) \to 0$ to conclude that

$$\lim_{n\to\infty} p_{52}(n) = 0$$

(iv) Result in course notes implies that $\mu_1 = \frac{1}{\pi_1} = \frac{1}{2/5} = \frac{5}{2}$.

MATH 37012 Solutions

2. (i)

$$P = \begin{bmatrix} p & q & 0 & 0 & \cdots \\ p & 0 & q & 0 & \cdots \\ 0 & p & 0 & q & \cdots \\ & & \ddots & & \ddots \end{bmatrix}$$

Then $\underline{\pi} P = \underline{\pi}$ gives

$$\begin{aligned} p\pi_0 + p\pi_1 &= \pi_0 \\ q\pi_0 + p\pi_2 &= \pi_1 \\ q\pi_1 + p\pi_3 &= \pi_2 \\ &\vdots \end{aligned} \qquad \begin{aligned} \pi_1 &= \frac{q}{p}\pi_0 \\ \pi_2 &= \left(\frac{q}{p}\right)^2\pi_0 \\ &\vdots \end{aligned}$$

Solving iteratively, or otherwise (no need to prove) gives us

$$\pi_k = \left(\frac{q}{p}\right)^k \pi_0 \qquad k \geq 1.$$

Then $1 = \sum_0^\infty \pi_k = \pi_0 \sum_0^\infty \left(\frac{q}{p}\right)^k$ has a non-negative

solution $\Leftrightarrow \frac{q}{p} < 1 \Leftrightarrow p > 1/2$, which is the range

of p for which a s.d. exists. In this case,

$$1 = \pi_0 \left(\frac{1}{1 - q/p}\right) \Rightarrow \pi_0 = (1 - q/p) \text{ and}$$

$$\pi_k = (1 - q/p)(q/p)^k \qquad k \geq 0.$$

(ii) Observe the Markov chain is irreducible & aperiodic. From course notes: no s.d. implies either all states transient or all null-recurrent. A s.d. exists then it is unique and $\mu_k = \frac{1}{\pi_k} < \infty$ and all states are positive recurrent. Therefore, all states are positive recurrent $\Leftrightarrow p > \frac{1}{2}$. In which case $\mu_0 = \frac{1}{1 - q/p}$.

2. (cont.) MATH 37012 Solutions

(iii)

$$\underline{\pi}\,\tfrac{1}{2}(I+P) = \tfrac{1}{2}\underline{\pi}I + \tfrac{1}{2}\underline{\pi}P = \frac{\underline{\pi}}{2} + \frac{\underline{\pi}}{2} = \underline{\pi}$$

Therefore $\underline{\pi}$ is also a s.d. for $\frac{1}{2}(I+P)$.

(iv) Using the hint, we see that this M.C. has transition matrix $\frac{1}{2}(I+P)$ where P is the transition matrix of $(X_n)_{n\geq 0}$ above, with $p=\frac{2}{3}$, $q=\frac{1}{3}$. Hence, from part (i), the s.d. is

$$\pi_k = \left(1-\frac{1/3}{2/3}\right)\left(\frac{1/3}{2/3}\right)^k = \left(\frac{1}{2}\right)^{k+1} \qquad k\geq 0.$$

(v) $\mu_0 = \frac{1}{\pi_0} = \frac{1}{1/2} = 2.$ Let T denote time of first visit to 0 from 1.

Then, starting at 0 and conditioning on first transition,

$$\mu_0 = \tfrac{1}{6}(E[T]+1) + \tfrac{5}{6}.1$$

$$\Rightarrow 12 = E[T]+1+5 \quad \Rightarrow E[T] = 6$$

MATH 3B012 Solutions

3. a)

(i) At the instant of entering state $i \geq 2$, there are the 2 new individuals plus $i-2$ partial lifetimes. By the l.o.m property, the time until the next birth is the minimum of i exponential (λ) r.v.s.

As they are independent, this has the exponential, parameter λi, distribution.

By general B.D process theory from course notes,

$$q_{i,i+1} = \lambda i \quad , \quad q_{i,i} = -\lambda i \qquad i \geq 1$$

and

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & \\ 0 & -2\lambda & 2\lambda & 0 & \cdots \\ 0 & 0 & -3\lambda & 3\lambda & 0 \cdots \\ & & & & \end{bmatrix}$$

In matrix form, the forward equations are $P'(t) = P(t)Q$, and we see that the first row is

$$p_{1k}'(t) = (k-1)\lambda p_{1,k-1}(t) - k\lambda p_{1k}(t) \qquad k \geq 1$$

as required.

b) Observe this is a Birth-death process.

$A \longleftrightarrow B$

$C \qquad D$

(i) In state 0, the time until the first conversation begins is the minimum of 4 iid exp (λ) r.v.s.

Consequently it is exp(4λ)

In state 1, the calls from C to D and D to C are both thinned Poisson processes with parameters $\frac{\lambda}{3}$.

MATH 37012 Solutions

Therefore the time until the next connection is $\exp(\frac{2\lambda}{3})$.
The time until the first completion is $\exp(\mu)$

In state 2, the time until the first completion is the minimum of 2 $\overset{iid}{\exp}(\mu)$ r.v's and hence is $\exp(2\mu)$.

This gives us $Q = \begin{bmatrix} -4\lambda & 4\lambda & 0 \\ \mu & -(\mu+\frac{2\lambda}{3}) & \frac{2\lambda}{3} \\ 0 & 2\mu & -2\mu \end{bmatrix}$

The s.d. is given by $\underline{\pi}\, Q = \underline{0}$ which is the system including

$$-4\lambda\pi_0 + \mu\pi_1 = 0$$
$$\frac{2\lambda}{3}\pi_1 - 2\mu\pi_2 = 0$$

giving $\pi_1 = 4\frac{\lambda}{\mu}\pi_0$, $\pi_2 = \frac{1}{3}\frac{\lambda}{\mu}\pi_1$. Then $\pi_0+\pi_1+\pi_2=1$

gives us $\pi_0\left(1 + 4\frac{\lambda}{\mu} + \frac{4}{3}\left(\frac{\lambda}{\mu}\right)^2\right) = 1$. And the

unique (by irreducibility) s.d. is $\left(\pi_0, \frac{4\lambda}{\mu}\pi_0, \frac{4}{3}\left(\frac{\lambda}{\mu}\right)^2\pi_0\right)$

(ii) Following the $\exp(\mu + \frac{2\lambda}{3})$ holding time in state 1, the process jumps to 0 w.p $\frac{\mu}{\mu+\frac{2\lambda}{3}}$ or 2 w.p. $\frac{2\lambda/3}{\mu+2\lambda/3}$. Consequently

$$\mu_{12} = \frac{1}{\mu+2\lambda/3} + \mu_{02}\frac{\mu}{\mu+\frac{2\lambda}{3}}$$

Also $\mu_{02} = \frac{1}{4\lambda} + \mu_{12}$. Solving for μ_{02}

gives $\mu_{02}\left(1 - \frac{\mu}{\mu+\frac{2\lambda}{3}}\right) = \frac{1}{\mu+\frac{2\lambda}{3}} + \frac{1}{4\lambda} \Rightarrow \mu_{02} = \frac{3}{2\lambda} + \frac{3\mu+2\lambda}{8\lambda^2}$

MATH 37012 Solutions

4. a)

(i) $S = \{0, 1, 2, \ldots\}$ This is a Birth-death process with

$q_{i,i+1} = \lambda_i \equiv \lambda$ and, using independence and lack of memory

$q_{i,i-1}\ \mu_i = i\mu \qquad i \geq 1.$

$q_{ii} \equiv -(\lambda + i\mu)$

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & \ldots & \\ \mu & -(\mu+\lambda) & \lambda & 0 & \ldots \\ 0 & 2\mu & -(2\mu+\lambda) & \ldots & \end{bmatrix}$$

Consequently, using expression derived in notes

$$\pi_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \pi_0 = \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \pi_0 \qquad n \geq 1$$

Then $\sum \pi_n = 1$ gives

$$\pi_0 \left(\sum_0^\infty \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \right) = 1 \Rightarrow \pi_0 e^{\lambda/\mu} = 1$$

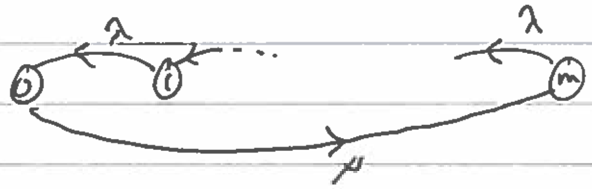
$\Rightarrow \pi_0 = e^{-\lambda/\mu}$. Therefore the s.d. is

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{e^{-\lambda/\mu}}{n!} \qquad n \geq 0$$ which we

recognise as Poisson, parameter λ/μ, which has mean λ/μ.

b)

(i)



$$Q = \begin{bmatrix} -\mu & 0 & \ldots & \ldots & 0 & \mu \\ \lambda & -\lambda & 0 & \ldots & \ldots & \\ 0 & \lambda & -\lambda & 0 & \ldots & \\ & & & & & \\ 0 & 0 & \ldots & \ldots & \lambda & -\lambda \end{bmatrix}$$

MATH 37012 Solutions

4. (cont.) $\underline{\pi} Q = 0$ gives us

$$-\mu\pi_0 + \lambda\pi_1 = 0$$
$$-\lambda\pi_1 + \lambda\pi_2 = 0$$
$$\vdots$$
$$-\lambda\pi_{m-1} + \lambda\pi_m = 0$$
$$\mu\pi_0 - \lambda\pi_m = 0$$

From these we get $\pi_1 = \frac{\mu}{\lambda}\pi_0$ $\quad \pi_1 = \pi_2 = \ldots = \pi_m$.

Then $\sum_1^m \pi_i = 1$ gives $\pi_0(1 + m\frac{\mu}{\lambda})$ $\Rightarrow$ $\pi_0 = \frac{1}{(1+\frac{m\mu}{\lambda})}$.

Hence the (unique) s.d. is $(\rho, \frac{\mu}{\lambda}\rho, \ldots, \frac{\mu}{\lambda}\rho)$
where $\rho = (1+\frac{m\mu}{\lambda})^{-1}$

(ii) This is just the sum of m exponential(λ) means, which is $\frac{m}{\lambda}$.

(iii) Conditioning on state at time of observation, this is

$$\sum_{i=1}^m \frac{i}{\lambda}\frac{\mu}{\lambda}\rho,$$

where we have used the lack of memory property.
The above is equal to

$$\frac{m(m+1)\mu}{2\lambda^2(1+\frac{m\mu}{\lambda})} = \frac{m(m+1)\mu}{2\lambda(\lambda+m\mu)}.$$

Two Hours

Statistical tables are attached

UNIVERSITY OF MANCHESTER

MEDICAL STATISTICS

31st May 2019

09:45-11:45

Please use SEPARATE booklets for your answers to each Section.

Answer **ALL** five questions in **SECTION A** (40 marks in total).

Answer **TWO** of the three questions in **SECTION B** (40 marks in total). If more than two questions from Section B are attempted, then credit will be given for the two best answers.

The total number of marks for the paper is 80.

University-approved calculators may be used.

Intended Learning Outcomes

At the end of this course unit a student will be able to:

1. Appraise the design, analysis and interpretation of the results of randomised controlled trials (RCTs), with specific reference to minimising bias.
2. Choose, with justification, the design, including calculation of the sample size, for a RCT with a continuous or binary outcome.
3. Analyse data from RCTs of various designs (including: parallel, crossover or cluster; superiority, non-inferiority or equivalence), including the meta-analysis of several RCTs, and interpret the findings in the relevant context.
4. Derive key mathematical expressions with applications to RCTs or meta-analyses of RCTs.

ALL BOOKWORK IS 'SEEN' (FROM EXERCISES OR NOTES) UNLESS STATED OTHERWISE

A1.

A randomised controlled trial of an **electronic** glucose monitor (**E**) against standard **finger**-prick glucose monitoring (**F**) in patients with Type 1 diabetes has been designed to use deterministic minimisation based on two baseline factors. The minimisation factors are trial centre (3 levels: Manchester; Birmingham, Cambridge) and HbA1c level (2 levels: 7.5%-9.0%; >9.0% to 11.0).

(Note: HbA1c is glycated haemoglobin, a measure of average glucose level over the past 2-3 months.)

The number of participants with each baseline factor level after 20 participants have been allocated to the intervention groups are given in the table below.

Baseline factor	Trial Centre						HbA1c level			
	Manchester		Birmingham		Cambridge		7.5% - 9.0%		>9.0% - 11.0%	
Treatment	E	F	E	F	E	F	E	F	E	F
Number of participants	4	2	3	3	3	5	5	6	5	4

(i) How many participants have been allocated to each treatment group?

ILO 1: Level Low

SOLUTION

Easiest to sum the numbers in Treatment E for the two HbA1c level, and likewise for Treatment F.

So, for E, 5+5=10 and for F, 6+4=10, i.e. 10 participants have been allocated to each treatment group.

[Calculation: 1 minute]

[1 mark]

(ii) The characteristics of the next participant (participant 21) entered into the trial is:

Trial Centre: Manchester

HbA1c level: 7.5%-9.0%.

Determine the intervention group to which this participant has to be allocated.

ILO 1: Level Medium

SOLUTION

Baseline factor	Trial Centre						HbA1c level				Sum: Manchester + 7.5%-9.0% for each treatment group	
	Manchester		Birmingham		Cambridge		**7.5% - 9.0%**		>9.0% - 11.0%			
Treatment	**E**	**F**	E	F	E	F	**E**	**F**	E	F	**E**	**F**
Number of participants	**4**	**2**	3	3	3	5	**5**	**6**	5	4	**9**	**8**

As the sum of the number of Manchester participants and the number of HbA1c level >9.0%-11.0% participants is greater for Treatment E than for Treatment F (9>8), participant 21 will be allocated to Treatment F.

[Calculation + reporting: 3 minutes]

[2 marks]

(iii) What type of bias has an increased risk when deterministic minimisation is used rather than stratified randomisation? How would the use of stochastic minimisation reduce this risk of bias?

ILO 1: Level Medium

SOLUTION

Selection (or allocation) bias is at increased risk when deterministic minimisation is used rather than stratified randomisation because the allocation is perfectly predictable if previous allocations are known (except in cases when the sums of previous allocations from the relevant participant baseline categories are equal form the treatment groups). The use of stochastic minimisation reduces this risk of bias as the allocation would not be known in this situation; it would only be known that there was an increased chance of allocation

[2 marks]

(iv) If stochastic randomisation is used with allocation probabilities 0.7 and 0.3, use the random number 0.60131 (from a uniform distribution on [0,1]) and a suitably-defined rule to allocate participant 21 to a treatment group.

ILO 1: Level: Low

SOLUTION

Using the allocation rule: Allocate to Treatment E if the random number is less than or equal to 0.3 and allocate to Treatment F if the random number is greater than 0.3, we would (still) allocate the participant to Treatment E as 0.60131 > 0.3.

[2 marks]

[Total 7 marks]

A2.

Van de Port (2012) reported testing for differences between treatment groups for 35 baseline variables, 4 of which had p-values less than 0.05. In the Discussion section of their paper they report some limitations of their trial, which included the text:

"Firstly, although the trial was well powered, including 250 patients with few drop-outs, we found significant baseline differences in favour of the circuit training group for a few secondary outcomes. All analyses, however, were adjusted for these covariates at baseline."

(i) Explain why performing significance tests for baseline differences between treatment groups is not recommended in randomised controlled trials.

ILO 1: Level Medium

SOLUTION

Performing significance tests for baseline differences between groups is not recommended because, assuming that randomisation has been performed correctly (and without allocation bias), any differences between treatment groups are due to chance. So the alternative hypothesis that the mean (or proportions) are not equal at baseline cannot be true as it is known that the mean (or proportions) *are* equal at baseline due to the allocation to the groups being random, meaning that the baseline distributions are the same for the two groups, and hence the population means *are* equal at baseline.

[2 marks]

(ii) When should the between-groups comparison of continuous outcome measures in randomized controlled trials be adjusted for baseline variables? Is it appropriate to use significance testing to select the baseline covariates to include in an adjusted analysis with the aim of reducing potential bias? (Justify your answer.)

ILO 1: Level Medium

SOLUTION

- Between-group comparison of continuous outcome measures should be adjusted for baseline variables when those baseline variables are likely to be prognostic of outcome.
- If such variables are unbalanced between the treatment groups, adjustment also reduces bias in the analysis. However, the choice as to which variables to adjust for should be made in advance of the data analysis (and documented in a Statistical Analysis Plan), so a significance test would be too late in the trial and, moreover, a p-value from a significance test is a function is not a measure of how large the difference actually is (which is what is important!).

[3 marks: 1 mark for first sub-part and 2 marks for second sub-part]

[Total 5 marks]

A3.

A randomised controlled trial is carried out to compare two doses of a new vaccine for the prevention of H1N1 influenza. The effectiveness of the vaccine is tested by measuring the immune response in the blood measured 9 days following vaccination (as this measures successful vaccination and protection against influenza). The results are summarised in the table below: the numbers in the table are the respective counts for the different combinations of Dose (15mg; 30mg) and Immune response (Yes; No), with the totals randomised to each of the dose levels.

Dose		15mg	30mg
Immune response	Yes	200	220
	No	50	30
	Total	250	250

(i) Estimate the odds ratio for the effectiveness of a 30mg dose as compared to a 15mg dose.

ILO 3: Level Low

SOLUTION

The required odds ratio estimate is given by $OR = \frac{220 \text{ x } 50}{30 \text{ x } 200} = 1.83$ (to 2 d.p.).

[Calculation: 1 minute]

[1 mark]

(ii) Calculate the 95% confidence interval of this odds ratio.

ILO 3: Level Medium

SOLUTION

The 95% confidence limits for the log odds ratio (30mg dose [T] relative to 15mg dose [C]) are given (approximately) by:

$\log_e(OR) \pm z_{0.025} SE(\log_e(OR))$ where $SE(\log_e(OR)) = \sqrt{\frac{1}{r_T} + \frac{1}{n_T - r_T} + \frac{1}{r_C} + \frac{1}{n_C - r_C}}$, where

$r_T, n_T - r_T, r_C$ and $n_C - r_C$ are the number of 'Yes' outcomes and 'No' outcomes for the treatment group and the number of 'Yes' outcomes and 'No' outcomes for the control group, respectively.

$\log_e(OR) = \log_e\left(\frac{220 \text{ x } 50}{30 \text{ x } 200}\right) = 0.6061, \ z_{0.025} = 1.9600$

$$SE(\log_e(OR)) = \sqrt{\frac{1}{220} + \frac{1}{30} + \frac{1}{200} + \frac{1}{50}} = \sqrt{0.06288} = 0.2508$$

So 95% confidence limits for $\log_e(OR)$ are 0.6061 - 1.9600 x 0.2508 = 0.1147

and 0.6061 + 1.9600 x 0.2508 = 1.0976.

Taking exponentials gives the 95% confidence Interval for the odds ratio as (1.12, 3.00) [limits to 2 d.p.]

[Calculation: 7 minutes]

[5 marks]

(iii) Interpret the results of this trial.

ILO 3: Level Low

SOLUTION

The trial suggests that the 30mg dose of the vaccine is more effective than the 15 mg dose as the odds ratio for immune response for a 30mg dose as compared to a 15mg dose equals 1.83 with 95% confidence interval 1.12 to 3.00.

[2 marks]

[Total 8 marks]

A4.

In a meta-analysis of *k* trials, suppose that $\hat{\theta}_i$ is an estimate of the treatment effect for the *i*th trial and let $Var[\hat{\theta}_i]$ be its sampling variance. The minimum variance estimate is defined by:

$$\hat{\theta}_{MV} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i}$$

where $w_i = {}^{1}\!/_{Var[\hat{\theta}_i]}$, and $Var[\hat{\theta}_{MV}] = \frac{1}{\sum_{i=1}^{k} \frac{1}{Var[\hat{\theta}_i]}}$.

The table below summarises the outcome of three trials of Neuromuscular Electrical Stimulation (NMES) versus no treatment for the condition of patellofemoral pain (PFP). The outcome measures was a Visual Analogue Scale (VAS) Pain score (range 0-10, with 0 representing no pain and 10 representing worst imaginable pain) during normal activities, assessed at the end of treatment. The treatment effect for each study (θ_i , i=1, 2, 3) is the difference in mean VAS pain for the two treatments (NMES group - No treatment group).

Study (Date of publication)	Difference $\hat{\theta}_i$	$Var[\hat{\theta}_i]$
Akarcali(2002)	1.12	0.3969
Bily (2008)	0.55	1.2769
Tunay (2003)	1.90	0.1296

(i) Compute the minimum variance estimate of the overall treatment effect, $\hat{\theta}_{MV}$, and determine its 95% confidence interval, stating any assumptions that you make.

ILO 3: Level Low

SOLUTION

	Difference $\hat{\theta}_i$	$Var[\hat{\theta}_i]$	$w_i = 1/Var[\hat{\theta}_i]$	$w_i\ \hat{\theta}_i$
	1.12	0.3969	2.519526	2.8219
	0.55	1.2769	0.783147	0.4307
	1.90	0.1296	7.716049	14.6605
Sum			11.0187	17.9131

So $\hat{\theta}_{MV} = \frac{\sum_{i=1}^{3} w_i \hat{\theta}_i}{\sum_{i=1}^{3} w_i} = \frac{17.9131}{11.0187} = 1.6257$ = **1.63 (to 2 d.p.)**

and $Var[\hat{\theta}_{MV}] = \frac{1}{\sum_{i=1}^{3} \frac{1}{Var[\hat{\theta}_i]}} = \frac{1}{11.0187} = 0.0908$

95% confidence limits for the overall treatment effect are given by:

$$\hat{\theta}_{MV} \pm z_{0.025}\sqrt{Var[\hat{\theta}_{MV}]}$$

i.e.

$$1.6257 \pm 1.9600\sqrt{0.0908}$$

giving 95% confidence limits of 1.0352 and 2.2161

So a 95% confidence interval for the treatment effect θ is **(1.04, 2.22)** (limits to 2 d.p.)

[Calculation 8 minutes]

[6 marks]

(ii) What can you conclude from the meta-analysis regarding the performance of NMES?

ILO 3: Level Low

SOLUTION

Pooled results from these three trials showed that NMES was associated with significantly lower pain scores at the end of treatment, when compared with a no-treatment control, with a mean treatment effect of 1.63 points in VAS pain during normal activity and a 95%CI of 1.04 to 2.22 points.

[2 marks]

(iii) It is known that the minimal clinically important difference for VAS pain is in the range 1.5 to 2.0. By comparing these values with the 95% confidence interval that you computed in (i), provide interpretation of the confidence interval in the context of clinical importance.

ILO 3: Level Medium

SOLUTION

Although the point estimate (1.63 points) of the overall treatment effect of NMES is just above the suggested lower limit (1.50 points) for a minimum clinically important difference (MCID) for VAS pain between the intervention groups, the *true* benefit (θ) may not be clinically important since the absolute value (1.04) for the lower 95% confidence limit was less than this suggested lower limit for the MCID.

[2 marks]

[Total 10 marks]

A5.

Consider a randomized controlled trial. Suppose the patient population can be divided into three latent sub-groups as follows:

- *Compliers*: patients who will comply with the allocated treatment;
- *Always control treatment*: patients who will receive control treatment regardless of allocation;
- *Always new treatment*: patients who will receive the new treatment regardless of allocation.

This assumes that there are no defiers, namely patients who will always receive the opposite of the treatment to which they are randomized. Assuming that the proportion and characteristics of compliers, *always control treatment*, *always new treatment* is the same in both arms and that randomization can only affect the outcome through the receipt of treatment, show that:

(i) An intention-to-treat estimate of the treatment effect is biased towards the null hypothesis of no treatment effect.

ILO 4: Level Medium

SOLUTION

Table of expected means under assumptions of model

	Type	Control Group	New Treatment Group	Proportion In Latent Class
As Randomized	**A**	μ	$\mu + \tau$	$\theta_A = 1 - \theta_B - \theta_C$
Always Control	**B**	$\mu + \gamma_B$	$\mu + \gamma_B$	θ_B
Always New Treatment	**C**	$\mu + \gamma_C + \tau$	$\mu + \gamma_C + \tau$	θ_C

For the Intention-to-Treat Estimate

$$\tau_{\mathrm{ITT}} = \left[\theta_A(\mu + \tau) + \theta_B(\mu + \gamma_B) + \theta_C(\mu + \gamma_C + \tau)\right] - \left[\theta_A\mu + \theta_B(\mu + \gamma_B) + \theta_C(\mu + \gamma_C + \tau)\right]$$
$$= \theta_A \tau$$

as the second and third terms in each bracket cancel.

Hence, as $0 \le \theta_A \le 1$, $|\hat{\tau}_{ITT}| \le |\tau|$ which means $\hat{\tau}_{ITT}$ is biased towards zero if $\theta_A < 1$, i.e. if some participants do not comply with treatment.

[Bookwork]

[5 marks]

(ii) A per-protocol estimate of the treatment effect may be biased either towards or away from the null hypothesis of no treatment effect.

ILO 4: Level High

SOLUTION

For the Per-Protocol Estimate

$$\tau_{PP} = \left[\frac{\theta_A(\mu + \tau) + \theta_C(\mu + \gamma_C + \tau)}{\theta_A + \theta_C}\right] - \left[\frac{\theta_A\mu + \theta_B(\mu + \gamma_B)}{\theta_A + \theta_B}\right]$$

$$= \left[\frac{(\theta_A + \theta_C)\mu + \theta_C\gamma_C + (\theta_A + \theta_C)\tau}{\theta_A + \theta_C}\right] - \left[\frac{(\theta_A + \theta_B)\mu + \theta_B\gamma_B}{\theta_A + \theta_B}\right]$$

$$= \tau + \mu + \left[\frac{\theta_C\gamma_C}{\theta_A + \theta_C}\right] - \mu - \left[\frac{\theta_B\gamma_B}{\theta_A + \theta_B}\right]$$

$$= \tau + \left[\frac{\theta_C\gamma_C}{1 - \theta_B - \theta_C + \theta_C}\right] - \left[\frac{\theta_B\gamma_B}{1 - \theta_B - \theta_C + \theta_C}\right]$$

$$= \tau + \left(\left[\frac{\theta_C\gamma_C}{1 - \theta_B}\right] - \left[\frac{\theta_B\gamma_B}{1 - \theta_B}\right]\right)$$

So τ_{PP} is biased by a term involving γ_B and γ_C. Since γ_B and γ_C can each be either positive or negative τ_{PP} may be biased either towards or away from zero.

[Bookwork]

[5 marks]

[Total 10 marks]

B1.

For an *AB*/*BA* crossover trial, a standard model for a continuous outcome y_{ij} for the i^{th} participant in the j^{th} period is:

$y_{i1} = \mu + \xi_i + \varepsilon_{i1}$ for a participant in sequence *AB* in period 1

$y_{i2} = \mu + \tau + \phi + \xi_i + \varepsilon_{i2}$ for a participant in sequence *AB* in period 2

$y_{i1} = \mu + \tau + \xi_i + \varepsilon_{i1}$ for a participant in sequence *BA* in period 1

$y_{i2} = \mu + \phi + \xi_i + \varepsilon_{i2}$ for a participant in sequence *BA* in period 2

where μ is the mean for the sequence *BA* in period 1, τ is the treatment effect of A compared to B, and ϕ is the period effect, with ξ_i and ε_{ij} being independent random variables with $\xi_i \sim N[0, \sigma_B^2]$ and $\varepsilon_{ij} \sim N[0, \sigma_\varepsilon^2]$. Defining $d_i = y_{i2} - y_{i1}$, let $\bar{d}_{AB}, \bar{d}_{BA}, \mu_{AB}^d$ and μ_{BA}^d be the sample and population means of d_i for sequences *AB* and *BA* respectively.

(i) Show that $\hat{\tau} = \frac{\bar{d}_{AB} - \bar{d}_{BA}}{2}$ is an unbiased estimator of τ, that is

$$E[\hat{\tau}] = E\left[\frac{\bar{d}_{AB} - \bar{d}_{BA}}{2}\right] = \tau .$$

Hence also show that a test of the null hypothesis $H_0{:}\ \mu_{BA}^d = \mu_{AB}^d$ is the same as a test of the null hypothesis of no treatment effect $H_0{:}\ \tau = 0$.

ILO 4: Level Low

SOLUTION

Substitution from the model given in the question gives:

$$E[d_i \mid i \in AB] = E\big[\big(\mu + \tau + \phi + \xi_i + \varepsilon_{i2}\big) - \big(\mu + \xi_i + \varepsilon_{i1}\big)\big]$$

The terms ξ_i on the RHS cancel so $E[d_i \mid i \in AB] = E[(\tau + \phi + \varepsilon_{i2}) - (\varepsilon_{i1})]$.

Since $E[\varepsilon_{ij}] = 0$, $E[d_i \mid i \in AB] = \tau + \phi$.

Hence $\mu^d_{AB} = E\left[\bar{d}_{AB}\right] = E\frac{\sum_{i\in AB} d_i}{n_{AB}} = \frac{n_{AB}E[d_i |\, i \in AB]}{n_{AB}} = \tau + \phi$.

Similarly, $\mu^d_{BA} = E[d_i \mid i \in BA] = E[(\phi + \varepsilon_{i2}) - (\tau + \varepsilon_{i1})\,] = \phi - \tau$.

And $E\left[\bar{d}_{BA}\right] = E\left[\frac{\sum_{i\in BA} d_i}{n_{BA}}\right] = \frac{n_{BA}E[d_i |\, i \in BA]}{n_{BA}} = \phi - \tau$.

Hence

$$\frac{\mu^d_{AB} - \mu^d_{BA}}{2} = E\left[\frac{\bar{d}_{AB} - \bar{d}_{BA}}{2}\right] = \frac{E\left[\bar{d}_{AB}\right] - E\left[\bar{d}_{BA}\right]}{2} = \frac{(\tau + \phi) - (\phi - \tau)}{2} = \tau$$

Likewise, a test of $H_o\colon \frac{\mu^d_{AB} - \mu^d_{BA}}{2} = 0$ or, equivalently, a test of $H_o\colon \mu^d_{AB} - \mu^d_{BA} = 0$ is the same as a test of the null hypothesis of no treatment effect $H_0\colon \tau = 0$.

[Bookwork]

[4 marks]

(ii) Show that the variance of $\hat{\tau}$ is given by

$$Var[\hat{\tau}] = Var\left[\frac{\bar{d}_{AB} - \bar{d}_{BA}}{2}\right] = \frac{\sigma^2_\varepsilon}{2}\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right).$$

(You may assume the general result $Var(\bar{x}) = \frac{\sigma^2}{n}$ where $Var(x) = \sigma^2$, and n is the sample size.)

ILO 4: Level Medium

SOLUTION

Define $Var[d_i |\, i \in AB] = \sigma^2_{d_{AB}}$ and $Var[d_i |\, i \in BA] = \sigma^2_{d_{BA}}$

From the model for a cross-over trial given in the question, for sequence AB

$$\sigma^2_{d_{AB}} = Var[d_i] = Var[y_{2i} - y_{1i}] = Var\left[\left(\mu + \phi + \xi_i + \varepsilon_{i2}\right) - \left(\mu + \tau + \xi_i + \varepsilon_{i1}\right)\right]$$

$$= Var[\varepsilon_{i2} - \varepsilon_{i1}] = Var[\varepsilon_{i1}] + Var[\varepsilon_{i2}] = 2\sigma_\varepsilon^2 \quad \text{since} \quad Cov[\varepsilon_{i1}, \varepsilon_{i2}] = 0$$

Similarly, for sequence BA, $\sigma_{d_{BA}}^2 = 2\sigma_\varepsilon^2$

Hence, the variances of the two sequences are the same for this model, that is:

$$\sigma_{d_{AB}}^2 = \sigma_{d_{BA}}^2 = \sigma_d^2 = 2\sigma_\varepsilon^2$$

Now $\quad \mathrm{Var}[\hat{\tau}] = \mathrm{Var}\left[\frac{\bar{d}_{AB}-\bar{d}_{BA}}{2}\right] = \frac{1}{4}Var\left[\bar{d}_{AB} - \bar{d}_{BA}\right]$

$$= \frac{1}{4}\left[Var\left(\bar{d}_{AB}\right) + Var\left(\bar{d}_{BA}\right)\right] \quad \text{as} \quad Cov\left[\bar{d}_{AB}, \bar{d}_{AB}\right] = 0$$

But $Var\left(\bar{d}_{AB}\right) = \frac{\sigma_d^2}{n_{AB}}$ and, likewise, $Var\left(\bar{d}_{BA}\right) = \frac{\sigma_d^2}{n_{BA}}$

So $\quad \mathrm{Var}[\hat{\tau}] = \frac{1}{4}\left[\frac{\sigma_d^2}{n_{AB}} + \frac{\sigma_d^2}{n_{BA}}\right] = \frac{1}{4}\left[\sigma_d^2\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right)\right] = \frac{2\sigma_\varepsilon^2}{4}\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right)$

$$= \frac{\sigma_\varepsilon^2}{2}\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right), \quad \text{as required.}$$

[Bookwork]

[5 marks]

(iii) Given that the variance of the estimator based on the first period observations from the two groups, which is equivalent to a parallel-design trial treatment effect estimator, is given by:

$$(\sigma_B^2 + \sigma_\varepsilon^2)\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right),$$

what does the result given in (iii) suggest is an advantage of using an *AB*/*BA* crossover design rather than a parallel-design trial? State the condition when this advantage is at its greatest.

ILO 1: Level Medium

SOLUTION

The ratio of the respective variances is $\frac{\frac{\sigma_\varepsilon^2}{2}}{(\sigma_B^2+\sigma_\varepsilon^2)} = \frac{\sigma_\varepsilon^2}{2(\sigma_B^2+\sigma_\varepsilon^2)}$.

So an advantage of using an *AB*/*BA* crossover design rather than a parallel-design trial is that the variance of the estimator is smaller by a factor of at least 2. The advantage is at its greatest when σ_B^2 is large relative to σ_ε^2.

[3 marks]

(iv) An AB/BA crossover trial was designed to compare two treatments for hypertension (high blood pressure): Drug A and Drug B. Thirty-eight participants, all suffering from hypertension, were randomised: 18 to the sequence Drug A then Drug B, and 20 to the sequence Drug B then Drug A. A measurement of the primary outcome measure, diastolic blood pressure (in mm Hg), was taken at the end of each treatment period and are summarised in the table below. Test the hypothesis H_0: $\tau = 0$ versus H_1: $\tau \neq 0$ using a 5% significance level. Also compute and briefly interpret the 95% confidence interval for τ.

	Period 1		*Period 2*		*Period 2 – Period 1*		
Sequence	*mean*	*s.d.*	*mean*	*s.d.*	*Mean*	*s.d*	*N*
AB	84.50	7.01	83.39	6.57	-1.11	2.47	18
BA	84.10	8.94	86.65	8.12	2.55	2.26	20

ILO 3: Level Low

SOLUTION

From part (i), the hypothesis H_0: $\tau = 0$ versus H_1: $\tau \neq 0$ can be tested using a two-sample t-test of the means of the differences H_0: $\mu_{BA}^d = \mu_{AB}^d$. The test statistic T_d is defined as:

$$T_d = \frac{\bar{d}_{AB} - \bar{d}_{BA}}{SE\left[\bar{d}_{AB} - \bar{d}_{BA}\right]}$$

where $\widehat{SE}[\bar{d}_{AB} - \bar{d}_{BA}] = s_d\sqrt{\frac{1}{n_{AB}} + \frac{1}{n_{BA}}}$ and $s_d = \sqrt{\frac{(n_{AB}-1)s^2_{d_{AB}}+(n_{BA}-1)s^2_{d_{BA}}}{n_{AB}+n_{BA}-2}}$

.

Substitution gives $s_d = \sqrt{\frac{17 \text{ x } 2.47^2+19 \text{ x } 2.26^2}{36}} = \sqrt{5.5767}$

Hence $\widehat{SE}[\bar{d}_{AB} - \bar{d}_{BA}] = \sqrt{5.5767\left(\frac{1}{18}+\frac{1}{20}\right)} = 0.7672.$

So $T_d = \frac{-1.11-2.55}{0.767} = -4.7704$

Under the null hypothesis this statistic will follow a t-distribution with 36 d.f.

From tables, $t_{0.025}(36) = 2.0281$. As $|T_d| = 4.7704 > 2.0281$ the null hypothesis H_0: $\tau = 0$ can be rejected in favour of H_1: $\tau \neq 0$ at the 5% significance level.

A 100(1-α)% confidence interval for the treatment effect τ is defined by its lower and upper limits:

$$\frac{1}{2}(\bar{d}_{AB} - \bar{d}_{BA}) \pm \frac{1}{2}t_{\alpha/2}(n_{AB}+n_{BA}-2)\,\widehat{SE}[\bar{d}_{AB} - \bar{d}_{BA}]$$

Therefore the 95% confidence limits for τ are

$$\frac{1}{2}(-3.66) \pm \frac{1}{2}2.0281 \text{ x } 0.7672$$

giving a 95% confidence interval for τ of (-2.61, -1.05) mm Hg (limits to 2 d.p.).

[Calculation: 10 minutes]

This can be briefly interpreted as follows: The true treatment effect is that the mean diastolic blood pressure is an average of between 1.05 and 2.61 mm Hg lower for patients given Drug A than for those given Drug B, with 95% confidence.

[8 marks]

[Total 20 marks]

B2.

A randomised controlled trial is planned to compare a treatment (A) with the current standard therapy (B). For an outcome measure *Y*, let $\overline{y_A}$, $\overline{y_B}$, μ_A, μ_B be the sample and population means of *Y* for each treatment. Let s and σ be the common within-group sample and population standard deviations of *Y*. Assume that the null hypothesis of no treatment effect will be tested by the statistic $T = \frac{\overline{y_A} - \overline{y_B}}{s\lambda}$, with $\lambda = \sqrt{1/n_A + 1/n_B}$, where n_A and n_B are the number of participants allocated to each treatment.

(i) Write down an expression for the power (1-β) of the test for a two-tailed alternative hypothesis to detect a treatment effect τ, stating the assumptions you make.

ILO 2: Level Medium

SOLUTION

The power of a trial with two groups of size n_T and n_C to detect an treatment effect of magnitude τ using a two-sample t-test with a two-sided significance level α is given by:

$$1 - \beta = 1 - \Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right)$$

where Φ is the cumulative distribution function of $N[0, 1]$ and $\lambda = \sqrt{\frac{1}{n_T} + \frac{1}{n_c}}$.

<u>Assumptions:</u>

1. Sample size is sufficiently large for normal distribution ($N[0, 1]$) to be a good approximation to the t-distribution on $n_T + n_C - 2$ degrees of freedom.
2. τ > 0 and $\frac{\tau}{\sigma\lambda}$ is sufficiently large for the probability in the left tail of the distribution to be negligible.

[Bookwork]

[4 marks]

(ii) Assuming that participants are allocated in the ratio of 1: k such that $n_B = k.n_A$, show that the total sample size required to give a power of (1-β) for a two-tailed α-sized test is:

$$n = \frac{(k+1)^2}{k}\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right) .$$

ILO 4: Level Medium

SOLUTION

Since $\Phi^{-1}(\beta) = -z_\beta$ it follows that $-z_\beta = z_{\alpha/2} - \frac{\tau}{\sigma\lambda}$, giving $\frac{\tau}{\sigma\lambda} = z_{\alpha/2} + z_\beta$.

If $n_T = kn_C$, then $\lambda = \sqrt{1/kn_C + 1/n_C}$

Therefore, $\frac{\tau}{\sigma}\sqrt{\frac{kn_C}{(k+1)}} = z_{\alpha/2} + z_\beta$.

Rearrangement gives $\frac{kn_C}{(k+1)} = \frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$

and further rearrangement gives $n_c = \left(\frac{k+1}{k}\right)\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$

Hence the total sample size is $n = N(k) = n_C + kn_C = \frac{(k+1)^2}{k}\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$, as required.

[Bookwork]

[5 marks]

(iii) Show that the minimum sample size is obtained by using an equal allocation ratio.

ILO 4: Level Medium

SOLUTION

$$N(k) = \frac{(k+1)^2}{k}\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$$

Now

$$N(k) = \left(\frac{k^2+2k+1}{k}\right)\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2 = \left(k+2+\frac{1}{k}\right)\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$$

So

$$N'(k) = \left(1-\frac{1}{k^2}\right)\frac{\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2 = 0 \text{ for min./max.}$$

Solving gives $k^2 = 1$ (as $z_{\alpha/2} + z_\beta > 0$) which gives $k = -1$ and $k = 1$ as solutions. However, $k = -1$ is not a valid solution given the context, so we are left with the solution $k = 1$.

Differentiating a second time gives $N''(k) = \frac{2}{k^3} = 2 > 0$ when $k = 1$, hence confirming that $k = 1$ does give a minimum for *n*.

So the minimum total sample size is when $k = 1$, that is when there is an equal allocation ratio.

[Bookwork - unseen]

[5 marks]

(iv) In a randomised controlled trial it is planned to allocate 240 participants to two treatments. The pooled within-group standard deviation is thought to be 12 units. Estimate the power of a study to detect a treatment effect of 5 units for two-tailed 0.05-sized test, if an allocation ratio of 1:2 (treatment: control) is used.

ILO 2: Level Low

SOLUTION

Power is given (approximately) by $1-\beta = 1-\Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right)$ where $\lambda = \sqrt{\frac{1}{n_T}+\frac{1}{n_c}}$.

If 240 participants are allocated in a ratio 1:2, then 1/3 of 240 = 80 participants will be allocated to the treatment group and 160 allocated to the control group. So $n_T = 80$ and $n_C = 160$.

Furthermore, $z_{\alpha/2} = 1.9600$ from tables and $\lambda = \sqrt{\frac{1}{80} + \frac{1}{160}} = \ 0.1369$,

So $\quad \frac{\tau}{\sigma\lambda} = \ \frac{5}{12 \text{ x } 0.1369} = \ 3.0429.$

Hence $\quad$ Power = $1 - \ \Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right) = 1 - \Phi(1.9600 - 3.0429) = 1 - \Phi(-1.0829).$

Now

$$\begin{aligned} \Phi(-1.0829) \ &\approx \ \Phi(-1.08) + 0.29 * \left(\Phi(-1.09) \ - \ \Phi(-1.08)\right) \\ &\approx (1 - 0.8599) + 0.29 * ((1 - 0.8621) - (1 - 0.8599)) \\ &\approx 1 - \ 0.8605 \end{aligned}$$

So $\quad$ Power = 1-(1-0.8605) = 0.8605, or 86.1%

[Calculation: 5 minutes]

[4 marks]

(v) In a randomised controlled trial, it is decided to allocate twice as many participants to the control group as to the treatment group, using blocked randomisation with a block size of 3. Briefly describe how to create a randomisation list to implement this randomisation scheme by using random numbers with a simple allocation rule.

[2 marks]

ILO 2: Level Medium

SOLUTION

Using this randomisation scheme, there are three possible blocks:

ABB, BAB, BBA,

each of which would be chosen with equal probability (1/3).

To create the randomisation list for this scheme, one could then randomly and repeatedly select, using a set of computer-generated pseudo-random numbers (uniformly distributed in the range [0,1]):

- the first block (ABB) if the pseudo-random number is <1/3 (0.33333);
- the second block (BAB) if the pseudo-random number is ≥1/3 but <2/3; and
- the third block (BBA) if the pseudo-random number is ≥2/3;

until the list is sufficiently long in order to randomise the necessary number of participants.

[2 marks]

[Total 20 marks]

B3.

In a parallel-group *non-inferiority* trial, a new treatment *T* is being compared with a control treatment *C* using a normally distributed outcome measure *Y*. Assume that large values of *Y* represent a worse outcome for the participant. Let $\overline{y_T}$, μ_T and n_T and $\overline{y_C}$, μ_C and n_C be the sample mean, population mean and the sample size for the new treatment and for the control treatment, respectively. Suppose σ is the population standard deviation of both treatments. The treatment effect is defined as $\tau = \mu_T - \mu_C$.

(i) Provide an example, with justification, of a situation where a non-inferiority trial rather than a superiority trial can be used.

ILO 2: Level Medium

SOLUTION

A non-inferiority trial can be used when the 'new' treatment is known (or can be shown to be) either less resource-intensive or have less side-effects than the current 'standard' treatment. So an example could be a trial where the new is cheaper or more efficient than the current standard treatment. Examples could include a nurse performing a procedure rather than a doctor (such as in the example for part (iv) where nurse-led telephone follow-up is used rather than doctor (consultant)-led clinic follow-up, where the use of a telephone rather than hospital premises is likely to add to the savings) or the use of physiotherapy rather than an operation (e.g. for frozen shoulder), or a drug or alternative therapy with less side-effects than the standard drug.

(Note: Any of these examples would be accepted for the marks.)

[2 marks]

(ii) Explain why a significance test of the hypothesis $H_0\colon \tau = 0$ versus $H_1\colon \tau > 0$ would not be appropriate in a *non-inferiority* trial.

ILO 1: Level Medium

SOLUTION

In order to demonstrate an effect we need to reject a null hypothesis so that we can accept an alternative hypothesis. Hence to demonstrate that a new treatment is non-inferior, we need to define a null hypothesis that the treatment is not non-inferior which can then be rejected in favour of an alternative hypothesis that the treatment is non-inferior to the control. A significance test of the hypothesis $H_0\colon \tau = 0$ versus $H_1\colon \tau > 0$ has a null hypothesis that the treatment is not effective (and hence non-inferior) whilst the alternative hypothesis is the hypothesis of inferiority (which the research is attempting to 'disprove'). So the hypotheses are not the 'correct way round', as the evidence from the data are always used to help support the alternative hypothesis (rather than the null hypothesis).

[3 marks]

(iii) For a non-inferiority trial, the null hypothesis $H_0\colon \tau \geq \tau_N$ is rejected if the upper limit for a $100(1-\alpha)\%$ one-sided confidence interval for τ is less than the limit of non-inferiority τ_N, for which the probability, under H_0, is given by:

$$\Pr[\text{Reject } H_0 |\ \tau\ ] = \Phi\left(\left(\frac{\tau_N}{\sigma\lambda} - z_\alpha\right) - \frac{\tau}{\sigma\lambda}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution and $\lambda = \sqrt{1/n_T + 1/n_C}$.

Given this, show that $\Pr[\text{Reject } H_0 | \tau]$ has a maximum under H_0 when $\tau = \tau_N$. Hence, show that this procedure has a Type I error probability less than or equal to α .

ILO 4: Level Medium

SOLUTION

$$\Pr[\text{Reject } H_0 \mid \tau] = \Pr[\hat{\tau} - z_\alpha \sigma\lambda > -\tau_N] = \Pr[\hat{\tau} > -\tau_N + z_\alpha \sigma\lambda] .$$

Since $\hat{\tau} \sim N[\tau, \sigma^2\lambda^2]$,

$$\Pr[\text{Reject } H_0 \mid \tau] = \Phi\left(\frac{\tau_N - z_\alpha\sigma\lambda + \tau}{\sigma\lambda}\right) : \text{ Equation (1)}$$

The maximum of this expression can be obtained by differentiation w.r.t. τ.

The derivative is:

$$\frac{d}{d\tau}\Pr[\text{Reject } H_0 \mid \tau] = \frac{1}{\sigma\lambda}\,\phi\left(\frac{\tau_N - z_\alpha\sigma\lambda + \tau}{\sigma\lambda}\right)$$

where ϕ is the standard normal density function.

Since ϕ>0 for finite values, it follows that $\frac{d}{dt}\Pr\,[\text{Reject } H_0 \mid \tau]$ is always positive and so $\Pr[\text{Reject } H_0 \mid \tau]$ is monotone increasing with τ as, under H_0 , $\tau \geq \tau_N$.

Hence, the Type 1 error rate has a maximum when $\tau = \tau_N$.

Setting $\tau = \tau_N$, in Equation (1).

$$\Pr[\text{Reject } H_0 \mid \tau] = \Phi\left(\frac{\tau_N - z_\alpha\sigma\lambda + \tau_N}{\sigma\lambda}\right) = \Phi(-z_\alpha) = \alpha .$$

Hence, the probability of a Type 1 error must be less than or equal to α.

[7 marks]

(iv) A randomised controlled non-inferiority trial was carried out to test whether *Nurse-led Telephone follow-up* (Telephone Group) was non-inferior to the current standard care of *Consultant-led Hospital follow-up* (Hospital Group) for women following treatment for cancer of the endometrium (the inner lining of the uterus). The primary outcome was measured on a continuous scale using a questionnaire-based measure of anxiety, with higher scores representing greater anxiety. Follow-up data were collected on 111 participants from the Telephone Group and 106 from the Hospital Group. The computer output from the statistical analysis is given below.

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [90% Conf. Interval]
---------+--------------------------------------------------------------------
telephon |     111    33.00901    1.045759    11.01775    31.27428    34.74374
hospital |     106     35.5283    1.265849    13.03271    33.42763    37.62897
---------+--------------------------------------------------------------------
combined |     217    34.23963    .8201926    12.08219    32.88472    35.59454
---------+--------------------------------------------------------------------
    diff |           -2.519293    1.635633               -5.221312    .1827265
------------------------------------------------------------------------------
    diff = mean(telephon) - mean(hospital)                        t =  -1.5403
Ho: diff = 0                                     degrees of freedom =      215

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0625         Pr(|T| > |t|) = 0.1250          Pr(T > t) = 0.9375
```

A difference of less than 3.5 points on the anxiety scale was considered by researchers to be an unimportant difference between treatments. Using the computer output given above, test whether the Nurse-Led Telephone follow-up was non-inferior to the Consultant-Led Clinic Hospital-based follow-up for the primary outcome measure. Interpret your findings in the context of the trial. Can you also conclude superiority of the Nurse-led Telephone follow-up on the primary outcome measure?

ILO 3: Level Medium

SOLUTION

The upper limit for a two-sided 90% confidence interval for the treatment effect τ is the same as the upper limit for a one-sided 95% confidence interval.

From the output this equals 0.704.

As 0.704 is less than 3.5, the limit of non-inferiority, the null hypothesis H_0: $\tau \geq \tau_N$ (treatment inferior) can be rejected in favour of the alternative hypothesis H_1: $\tau < \tau_N$ (treatment non-inferior).

[3 marks]

In the context of the trial, this can be interpreted as follows:

The Nurse-Led Telephone follow-up was significantly non-inferior to the Consultant-Led Clinic Hospital-based follow-up for anxiety at the 5% significance level, based on the upper confidence limit for the treatment effect, 0.70 being less than the non-inferiority limit of 3.5 points on the anxiety scale.

[2 marks]

It is not possible, however, to conclude that the Nurse-Led Telephone follow-up was significantly non-inferior to the Consultant-Led Clinic Hospital-based follow-up for anxiety at the 5% significance level. This would involve testing $H_0: \tau = 0$ against $H_1: \tau \neq 0$. The p-value for this test is given in the computer output as 0.125 > 0.05 so $H_0: \tau = 0$ cannot be rejected at the 5% significance level and so we cannot conclude that the Nurse-Led follow-up is superior to the Consultant-Led Clinic Hospital-based follow-up in terms of patient anxiety.

[3 marks]

[Total part (iv): 8 marks]

[Total 20 marks]

<u>**Total marks:**</u> **ILO 1: 18 (A:12; B:6); ILO2: 12 (A:0; B:12); ILO3: 34 (A:18; B:16); ILO4: 36 (A:10; B:26)**

END OF EXAMINATION PAPER SOLUTIONS

Solutions for MATH39032 Mathematical Modelling in Finance,

ILO1. Recognise the role that financial derivatives play in reducing risk.
ILO2 Construct payoff diagrams for standard options (and portfolios of options).
ILO3 Construct a PDE, using the concepts of stochastic calculus and hedging.
ILO4 Solve analytically the standard Black-Scholes equation.
ILO5 Use the Black-Scholes formulae to evaluate fair prices for European options.
ILO6 Extend the basic European option model (to include dividends and/or early exercise) and where possible to solve the resulting models analytically.

Question 1

Learning Outcome	**Solution**
ILO1	*Bookwork. ILO1 is tested at a low level.*

[correct answer 6 marks]

We set up two portfolios, A and B at $t=0$

$$\Pi_0^A = C_0 + X e^{-rT} \quad \text{and} \quad \Pi_0^B = P_0 + S_0$$

At expiry, $t=T$:

3

$$\Pi_T^A = \begin{cases} S - X + X = S, & S \geq X \\ X & S < X \end{cases}, \quad \Pi_T^B = \begin{cases} S & S \geq X \\ X - S + S = X & S < X \end{cases}$$

$$\Pi_T^A = \Pi_T^B \Rightarrow \Pi_t^A = \Pi_t^B \text{ by no arbitrage}$$

Therefore

$$C(S,t) + X e^{-r(T-t)} = P(S,t) + S$$

at $t=0$

3

$$C_0 + X e^{-rT} = P_0 + S_0$$

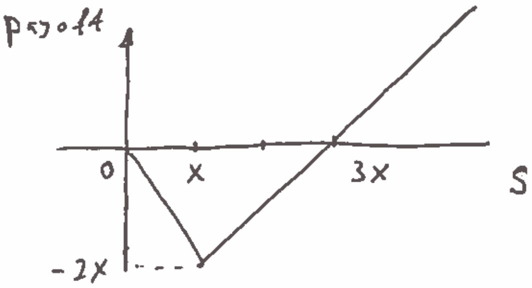
Question 2

Learning Outcome	Solution
ILO2	*Similar problem in examples sheets. ILO2 is tested at a low level.*

Payoff diagrams: [correct answer 12 marks]

(i) $\Pi = -2S + 3C$

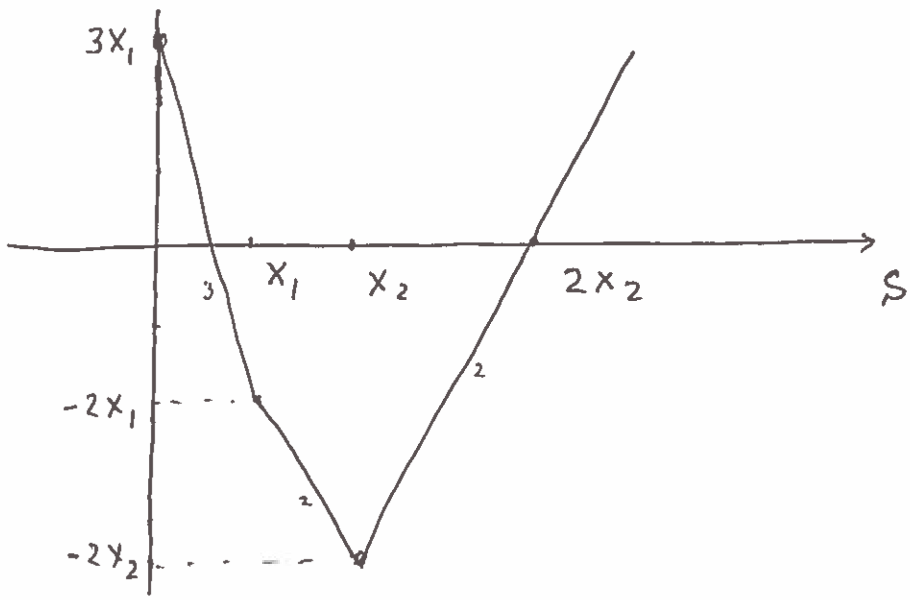
$$\Pi_T = \begin{cases} -2S & S < X \\ -2S + 3(S-X) = S - 3X & S \geq X \end{cases}$$



4

(ii) $\Pi = -2S + 3P(X_1) + 4C(X_2)$, $X_1 < X_2$

$$\Pi_T = \begin{cases} -2S + 3(X_1 - S) = 3X_1 - 5S & S \leq X_1 \\ -2S & X_1 < S < X_2 \\ -2S + 4(S - X_2) = 2S - 4X_2 & S \geq X_2 \end{cases}$$



8

Question 3

Learning Outcome	**Solution**
ILO3	*Similar problem in lecture notes. ILO3 is tested at a medium level.*

[correct answer 12 marks]

By using Itô's Lemma from the lecture notes (see formula (7)), we obtain

$$dV = \left[0.5 S \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2}(1+t)^2 S \frac{\partial^2 V}{\partial S^2} \right] dt$$

$$+ \; 0.5\, S \cdot \frac{\partial V}{\partial S} \Big] dW$$ 4

We set up a portfolio

$$\Pi = V - \Delta S \quad \text{for which}$$

$d\Pi = dV - \Delta dS$. Substituting in the expressions for dV and dS, we obtain

$$d\Pi = \left[\frac{\partial V}{\partial t} + \frac{1}{2}(1+t)^2 S \frac{\partial^2 V}{\partial S^2} \right] dt$$ 4

under the choice

$$\Delta = \frac{\partial V}{\partial S}$$

By no-arbitrage

$\frac{d\Pi}{\Pi} = r\,dt$, where $\Pi = V - \frac{\partial V}{\partial S} S$, where

r - risk-free interest rate

we obtain

$$\boxed{\frac{\partial V}{\partial t} + \frac{1}{2}(1+t)^2 S \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0}$$ 4

Question 4

Learning Outcome	**Solution**
ILO4	*Similar problem in lecture notes and examples sheets. ILO4 is tested at a medium level.*

[correct answer 20 marks]

Let us find partial derivatives for

$V(S,t) = Ae^{rT} S^{\beta}$:

(i) $\frac{\partial V}{\partial t} = rV$, $\frac{\partial V}{\partial S} = \frac{\beta}{S} V$, $\frac{\partial^2 V}{\partial S^2} = -\frac{\beta}{S^2} V + \frac{\beta}{S}\frac{\partial V}{\partial S} =$

$= -\frac{\beta V}{S^2} + \frac{\beta^2 V}{S^2}$ Substitution gives 4

$r\cancel{V} + \frac{1}{2}\sigma^2(\beta^2 - \beta)V + r\beta V - r\cancel{V} = 0$ $V \neq 0$

$\frac{1}{2}\sigma^2\beta^2 + \beta(r - \frac{1}{2}\sigma^2) = 0$

Two solutions : $\boxed{\beta = 0}$ and $\beta = 2\frac{\frac{\sigma^2}{2} - r^2}{\sigma^2}$

(ii) $\boxed{\beta = 1 - \frac{2r^2}{\sigma^2}}$ 6

(iii) By using chain rule , we find

$\frac{\partial V}{\partial S} = \frac{\partial V}{\partial z}\frac{dz}{dS} = -\frac{\partial V}{\partial z}\frac{1}{S}$, $\frac{\partial^2 V}{\partial z^2} = \frac{\partial V}{\partial z}\frac{1}{S^2} - \frac{\partial^2 V}{\partial z^2}\left(-\frac{1}{S}\right)\frac{1}{S} =$

$= \frac{1}{S^2}\frac{\partial V}{\partial z} + \frac{1}{S^2}\frac{\partial^2 V}{\partial z^2}$. Substitution gives 4

$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\left[\frac{1}{S^2}\frac{\partial V}{\partial z} + \frac{1}{S^2}\frac{\partial^2 V}{\partial z^2}\right] + rS\left[-\frac{\partial V}{\partial z}\frac{1}{S}\right] - rV = 0$

Finally

$$\boxed{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial z^2} + \left(\frac{1}{2}\sigma^2 - r\right)\frac{\partial V}{\partial z} - rV = 0}$$ 6

Question 5

Learning Outcome	**Solution**
ILO5	*Similar problem in examples sheets. ILO5 is tested at a high level.*

(i) $S_0 = 100$, $X = 100$, $T = 0.25$, $\sigma = 1$, $r = 0$ [correct answer 25 marks]

First we compute the values of d_1 and d_2:

$$d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} = \frac{1}{2}\sqrt{0.25} = 0.25, \quad d_2 = -0.25$$

$C_0 = 100\,N(0.25) - 100\,N(-0.25)$, where

$N(0.25) = 0.5987$ and $N(-0.25) = 0.4013$

$$C_0 = 59.85 - 40.13 = \boxed{19.72}$$

4

(ii) As $X \to 0$, we see that $d_1 \to \infty$ and $d_2 \to \infty$. Therefore, $N(d_1) \to 1$ and $N(d_2) \to 0$ from the formula, we have

$$\boxed{\lim_{X \to 0} c(S,t) = S}$$

8

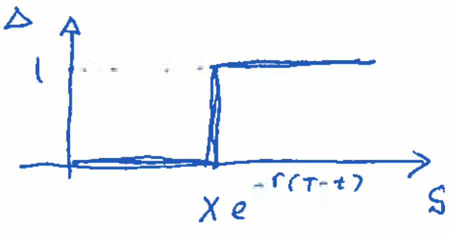
As $\sigma \to \infty$, we see that $d_1 \to \infty$ and $d_2 \to -\infty$

$N(d_1) \to 1$, $N(d_2) \to 0$ $\Rightarrow$

$$\boxed{\lim_{\sigma \to \infty} c(S,t) = S}$$

see next page for the limit $\sigma \to 0$

(iii) $$\Gamma = \frac{\partial^2 C}{\partial S^2} = \frac{\partial \Delta}{\partial S} = N'(d_1)\frac{\partial d_1}{\partial S} = \boxed{\frac{e^{-\frac{d_1}{2}}}{S\sigma\sqrt{2\pi(T-t)}}}$$



6

Question 5

Learning Outcome	Solution
ILO5	*Similar problem in examples sheets. ILO5 is tested at a high level.*

[correct answer 25 marks]

(iv) $\frac{\partial c}{\partial X} = SN'(d_1)\frac{\partial d_1}{\partial X} - X e^{-r(T-t)} N'(d_2)\frac{\partial d_2}{\partial X} - e^{-r(T-t)} N(d_2) =$

$[SN'(d_1) - X e^{-r(T-t)} N'(d_2)]\frac{\partial d_1}{\partial X} - e^{-r(T-t)} N(d_2)$

Let us show that $SN'(d_1) - X e^{-r(T-t)} N'(d_2) = 0$

and therefore $\boxed{f(t) = -e^{-r(T-t)}}$ 7

We obtain

$N'(d_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}}$ and $N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_2^2}{2}} =$

$= N'(d_1) e^{d_1 \sigma\sqrt{T-t} - \sigma^2 (T-t)/2} = N'(d_1) e^{\ln(\frac{S}{X}) + r(T-t)}$

$= N'(d_1) \frac{S}{X} e^{r(T-t)}$

Then

$SN'(d_1) - \cancel{X} e^{-r(T-t)} N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}} \left(S - \cancel{X} \cancel{e^{-r(T-t)}} \frac{S}{\cancel{X}} \cancel{e^{r(T-t)}}\right) = 0$

(ii) if $S - X e^{-r(T-t)} > 0$ or $\ln(\frac{S}{X}) + r(T-t) > 0$

as $\sigma \to 0$, we see that $d_1 \to \infty$, $d_2 \to \infty$

$\boxed{\lim_{\sigma\to 0} C(S,t) = S - X e^{-r(T-t)}}$

if $S - X e^{-r(T-t)} < 0$ or $\ln(\frac{S}{X}) + r(T-t) < 0$

$\boxed{\lim_{\sigma\to 0} C(S,t) = 0}$

Question 6

Learning Outcome	Solution
ILO6	*Similar problem in lecture notes and examples sheets. ILO6 is tested at a high level.*

[correct answer 25 marks]

Sergei Fedotov
March 2, 2019

(i)

Let us find derivatives:

$$\frac{\partial P}{\partial S} = \frac{\partial v}{\partial z}\frac{\partial z}{\partial S} = \frac{\partial v}{\partial z}\frac{1}{S}$$

$$\frac{\partial^2 P}{\partial S^2} = -\frac{1}{S^2}\frac{\partial v}{\partial z} + \frac{1}{S}\frac{\partial^2 v}{\partial z^2}\frac{\partial z}{\partial S} =$$

$$= -\frac{1}{S^2}\frac{\partial v}{\partial z} + \frac{1}{S^2}\frac{\partial^2 v}{\partial z^2}$$

$$\frac{\partial P}{\partial \tau} = -\frac{\partial v}{\partial \tau}$$ Substitution gives

$$-\frac{\partial v}{\partial \tau} + \frac{1}{2}\sigma^2\left[-\frac{\partial v}{\partial z} + \frac{\partial^2 v}{\partial z^2}\right] - rv = 0$$

$$\boxed{\frac{\partial v}{\partial \tau} = \frac{1}{2}\sigma^2\frac{\partial^2 v}{\partial z^2} - \frac{1}{2}\sigma^2\frac{\partial v}{\partial z} - rv}$$

6

Question 6

Learning Outcome	Solution
ILO6	*Similar problem in lecture notes and examples sheets. ILO6 is tested at a high level.*

[correct answer 25 marks]

Sergei Fedotov
March 2, 2019

(ii) Substitution $P = AS^{\beta}$ into

$$\frac{\partial P}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 P}{\partial S^2} - rP \quad \text{gives}$$

$\frac{1}{2}\sigma^2 \beta(\beta-1) - r = 0$. We obtain $\beta_{1,2} = \dfrac{\frac{1}{2}\sigma^2 \pm \sqrt{\frac{1}{4}\sigma^4 + \frac{4}{2}\sigma^2 r}}{2\,\frac{1}{2}\sigma^2}$

$$\boxed{\beta_{1,2} = 1 \pm \sqrt{1 + \frac{8r}{\sigma^2}}}$$

6

(iii) Two conditions: $P(S_f) = X - S_f$, $\frac{\partial P}{\partial S}(S_f) = -1$

(iii) since $P(S \to \infty) \to 0$, we have

$$\boxed{\beta = 1 - \sqrt{1 + \frac{8r}{\sigma^2}} < 0}$$

4

(iv) Let us find A and S_f from two equations:

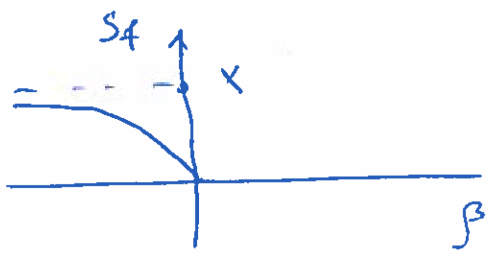
$A S_f^{\beta} = X - S_f$ and $\beta A S_f^{\beta-1} = -1$

Location of free boundary: $\boxed{S_f = \dfrac{\beta X}{\beta - 1}}$

Constant $A = -\dfrac{1}{\beta S_f^{\beta-1}} = \boxed{-\left(\dfrac{\beta X}{\beta-1}\right)^{1-\beta}\dfrac{1}{\beta}}$

9

(v)



As $D \to 0$, $\beta \to 0$, $S_f \to 0$

Solutions to exam MATH39512 Actuarial Models 2 2018

General remarks MATH39512 Actuarial Models 2 covers units 5 (on basic definitions and concepts in survival analysis; tested throughout this exam paper), 6 (on non-parametric and parametric estimation of survival times; see Questions 1 and 3) and 7 (on the Cox PH model; see Question 4) of the syllabus of CT4 of the Actuarial Profession. The material covered in this course unit goes beyond this particular part of the CT4 syllabus as other topics in survival analysis, like (left-)truncation, the log-rank test, frailty and applications of martingales in survival analysis are taught in this course unit as well. The individual learning outcomes (ILOs) associated with this course unit are

1. Determine whether a given censoring mechanism for the purpose of estimating survival times is independent or not.
2. Perform estimation of truncated and censored survival times using non-parametric, parametric and semi-parametric methods.
3. Carry out a variety of hypothesis tests (in particular, 2-sample, likelihood related and graphical tests) associated with the estimation procedures and interpret the outcomes.
4. Use martingales associated with counting processes for computing expectations and variances of estimators and test statistics of survival times.
5. Describe the frailty effect and how it can explain certain artefacts that might appear (like decreasing population hazard, crossover phenomenon) in a given proportional frailty model.

Regarding the originality level of the questions and how they are aligned with the ILOs and at what level (low/medium/high):

- Question 1 on non-parametric estimation: parts (a) and (b) are standard, part (c) is new while (d) is a variation of an exercise. This question is associated with ILO 2 with part (a) at low to medium level, (b) at low level, (c) at medium level and (d) at medium to high level.
- Question 2 on independent censoring: students have seen such questions before but not within this particular context. This question is associated with ILO 1 at medium level.
- Question 3 on parametric estimation: part (a) is very similar to an exercise but with the difference that now a derivation is required as well. This part is associated with ILO 2 at medium to high level. Regarding part (b), students have seen similar questions though typically within a non-parametric estimation setting rather than a parametric one as is the case here. Further subpart (ii) is unusually easy. Part (b)(i) is associated with ILO 3 at low to medium level while parts (ii) and (iii) are associated with ILO 4 at respectively low and high level.
- Question 4 on Cox PH model: fairly standard though the application is original and students typically find estimation with the Cox PH model difficult. Regarding part (d), students are used to being asked to carry out a hypothesis test rather than to describe one. Parts (a)-(c) are associated with ILO 2 at medium level and part (d) is associated with ILO 3 at medium level.
- Question 5 on frailty: this particular proportional frailty model has not been seen before. Regarding parts (b) and (c), students have seen similar questions in the exercises but the context is original and the style is somewhat different here. This question is associated with ILO 5 at medium level.

Answer to 1

(a) We first find the values for τ_i (the ith ordered observed genuine failure time), d_i (the observed number of failures at τ_i) and r_i (observed number of individuals at risk of failing just before τ_i). The values are given in the table below. Then we can compute the Kaplan-Meier estimate $\widehat{S}(t)$ at the failure times τ_i using $\widehat{S}(\tau_1) = (1 - d_1/r_1)$ and the recursion, $\widehat{S}(\tau_i) = \widehat{S}(\tau_{i-1})(1 - d_i/r_i)$, $i = 2, \ldots, k$, where k denotes the total number of distinct genuine failure times. This leads to the last column of the table below. Note that as $\widehat{S}(t)$ is a right-continuous step function, the value of $\widehat{S}(t)$ for $t \in [\tau_i, \tau_{i+1})$, $i = 1, \ldots, k-1$, is given by $\widehat{S}(t) = \widehat{S}(\tau_i)$, whereas $\widehat{S}(t) = 1$ for $t \in [0, \tau_1)$ and $\widehat{S}(t) = \widehat{S}(\tau_k)$ for $t \geq \tau_k$. The plot of the Kaplan-Meier estimate is provided in Figure 1.

i	τ_i	d_i	r_i	$\widehat{S}(\tau_i)$
1	3	1	3	0.667
2	10	1	10	0.600
3	11	1	8	0.525
4	12	2	7	0.375
5	16	1	5	0.300
6	25	1	1	0.000

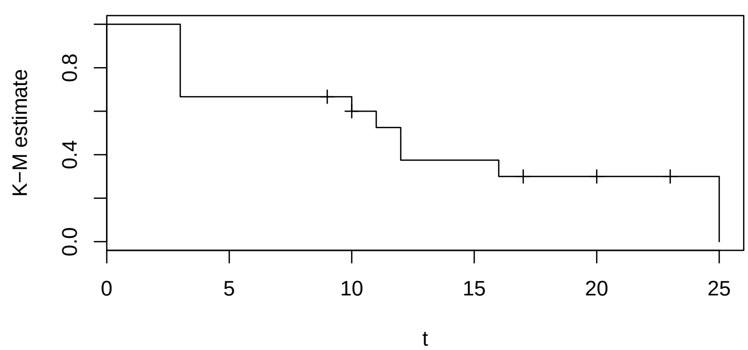


Figure 1: Plot of the Kaplan-Meier estimate.

[Mark distribution: 3 for correct definition with explanation of notation, 3 for correct table, 2 for correct final answer and plot.]

(b) With T the random variable representing the failure time of an individual we need to estimate $\mathbb{P}(T \leq 20) = 1 - \mathbb{P}(T > 20)$. Using the Kaplan-Meier estimate, we can estimate this quantity by

$$1 - \widehat{S}(20) = 1 - 0.300 = 0.700,$$

where the value of $\widehat{S}(20)$ can be estimated from Figure 1 (in combination with the table above it).

(c) If the survival data corresponding to individual 1 is $(0, 3]$ instead of $(1, 3]$, then there are more individuals at risk at the beginning which leads to an increase of the Kaplan-Meier estimate at that time, which in turn leads to an increase of the Kaplan-Meier estimate at any time due to its product structure. Simply said, the exposed to risk increases while the number of failures remains the same and so the estimated survival probability increases. This means that the estimate of the probability in part (b) goes down.

(d) We want to estimate the quantity

$$\mathbb{E}\left[T|T\leq 15\right]=\frac{\mathbb{E}\left[T\mathbf{1}_{\{T\leq 15\}}\right]}{\mathbb{P}(T\leq 15)}=\frac{\int_0^{15}\mathbb{P}(T>t)\mathrm{d}t-15\mathbb{P}(T>15)}{\mathbb{P}(T\leq 15)}=\frac{\int_0^{15}\mathbb{P}(T>t)\mathrm{d}t-15\mathbb{P}(T>15)}{1-\mathbb{P}(T>15)}.$$

We can estimate this quantity, using the Kaplan-Meier estimate, via

$$\begin{aligned}
&\frac{\int_0^{15}\widehat{S}(t)\mathrm{d}t-15\widehat{S}(15)}{1-\widehat{S}(15)}\\
=&\frac{1(3-0)+0.667(10-3)+0.6(11-10)+0.525(12-11)+0.375(15-12)-15\cdot 0.375}{1-0.375}\\
=&\frac{4.291667}{0.625}\\
=&6.867.
\end{aligned}$$

Answer to 2

A censoring mechanism is called independent if under this mechanism the hazard rate of an individual who is at risk at time t (i.e. has not experienced the event in question (which in this case is dying due to Alzheimer's disease) and has not been censored) is the same as that of an individual who is at risk at time t in the hypothetical situation where there is no such censoring. So with independent censoring the individuals who remain at risk are 'representative' for the sample of patients that we would have had, if there had been no censoring.

In this example, the censoring is independent since there is no reason why the hazard rate of a person at risk would alter when the study would run its full course in comparison to the current case where it is stopped early because of too many patients dying. This kind of censoring (namely type II censoring) is actually quite common, especially in the area of reliability and is known to be independent.

As a side remark, note that although the censoring, i.e. stopping the treatment early, might well save some lives, this affects the hazard rate of dying due to Alzheimer's disease after censoring occurs, but for verifying whether or not the censoring is independent one needs to look at the hazard rate of an individual before censoring occurs and compare this with the (hypothetical) case where the censoring mechanism is removed. Since an individual who is still at risk either receives the treatment in both situations or does not receive the treatment in both situations, the hazard rates in the aforementioned two situations are the same (and so the censoring is independent).

[Mark distribution: 2 for correct definition, 3 for correct explanation.]

Answer to 3

(a) Denote by T the generic survival time with hazard function $\mu(t)$. Since the hazard function is $\mu(t)=\alpha\lambda^\alpha t^{\alpha-1}$, the cumulative hazard function equals

$$A(t)=\int_0^t\mu(s)\mathrm{d}s=\int_0^t\alpha\lambda^\alpha s^{\alpha-1}\mathrm{d}s=\lambda^\alpha s^\alpha\big|_{s=0}^t=\lambda^\alpha t^\alpha$$

and so the survival function is

$$S(t)=\exp\left(-A(t)\right)=\exp\left(-\lambda^\alpha t^\alpha\right).$$

Hence the pdf of T is

$$f(t)=\mu(t)S(t)=\alpha\lambda^\alpha t^{\alpha-1}\exp\left(-\lambda^\alpha t^\alpha\right).$$

Let $L_i(\alpha,\beta)$ be the likelihood of (α,β) given the data of individual i only, where the individuals are labelled 1 to 5 from left to right. For a censored observation t_i+, the likelihood by the independence of the censoring and the survival time is $L_i(\alpha,\beta) = \Pr(T > t_i)c_i = S(t_i)c_i$, where c_i is some unimportant constant (i.e. does not depend on α or β). Similarly, for an uncensored observation t_i, the likelihood is $L_i(\alpha,\beta) = \mathrm{Lik}(T = t_i)c_i = f(t_i)c_i$, where again c_i is some unimportant constant. Then given that individuals 1, 2 and 4 fail and the others are censored and that all individuals are independent, the likelihood of all individuals combined is,

$$L(\alpha,\beta) = \prod_{i=1}^{5} L_i(\alpha,\beta) = Cf(2)f(5)f(6)S(5)S(7) = C\mu(2)\mu(5)\mu(6)S(2)S(5)^2S(6)S(7),$$

where C is an unimportant constant and for the last equality we used $f(t) = \mu(t)S(t)$. Hence, recalling that $S(t) = \mathrm{e}^{-A(t)}$, the log-likelihood is

$$\ell(\alpha,\beta) = \log(L(\alpha,\beta)) = \log(C) + \log(\mu(2)) + \log(\mu(5)) + \log(\mu(6)) - A(2) - 2A(5) - A(6) - A(7),$$

Since $A(t) = \lambda^\alpha t^\alpha$ and $\log(\mu(t)) = \log(\alpha\lambda^\alpha) + (\alpha-1)\log(t)$, we get in the end,

$$\ell(\alpha,\beta) = \log(C) + 3\log(\alpha\lambda^\alpha) + (\alpha-1)[\log(2) + \log(5) + \log(6)] - \lambda^\alpha[2^\alpha + 2\cdot 5^\alpha + 6^\alpha + 7^\alpha].$$

[Mark distribution: 3 for correct general likelihood formula in terms of hazard, cumulative hazard and/or survival function, 2 for its derivation, 3 for inserting the data into the formula correctly.]

(b) (i) We have $N_\infty = 3$ and

$$R_s = \begin{cases} 5 & \text{if } s \in [0,2], \\ 4 & \text{if } s \in (2,5], \\ 2 & \text{if } s \in (5,6], \\ 1 & \text{if } s \in (6,7], \\ 0 & \text{if } s \in (7,\infty), \end{cases}$$

which implies $\int_0^\infty R_s \mathrm{d}s = 5\cdot 2 + 4\cdot 3 + 2\cdot 1 + 1\cdot 1 = 25$. Hence

$$\left|\frac{N_\infty - \lambda\int_0^\infty R_s\mathrm{d}s}{\sqrt{\lambda\int_0^\infty R_s\mathrm{d}s}}\right| = \left|\frac{3 - 0.20\cdot 25}{\sqrt{0.20\cdot 25}}\right| = \frac{2}{\sqrt{5}} = 0.89.$$

Since $0.89 < 1.96 = z_{0.025}$ we do not reject the null hypothesis at significance level 0.05 with this 1-sample test. We conclude that according to the 1-sample test there is no evidence at the 5% significance level that the hazard rate is not a constant equal to 0.20.

(ii) Under H_0, $\mu(t) = \lambda$ for all $t \geq 0$, so under H_0, $Z = M$. So since M is a martingale, then so is Z under H_0.

(iii) Since martingales have constant expectations and $Z_0 = N_0 - 0 = 0$ (since with probability 1 no individual dies at time 0), we have for any $t \geq 0$, $\mathbb{E}[Z_t] = \mathbb{E}[Z_0] = 0$.
For the variance, we use a result from the notes that the variance of a 0-mean martingale is equal to the expectation of its square bracket process, i.e. $\mathrm{Var}(M_t) = \mathbb{E}[[M]_t]$, where the square bracket process $[M] = \{[M]_t : t \geq 0\}$ is defined as $[M]_t = \sum_{0<s\leq t}(\Delta M_s)^2$ with $\Delta M_s := M_s - \lim_{u\uparrow s} M_s$. Since under H_0, Z is a martingale, we thus deduce, for any $t \geq 0$,

$$\mathrm{Var}(Z_t) = \mathbb{E}[[Z]_t] = \mathbb{E}\left[\sum_{0<s\leq t}(\Delta Z_s)^2\right] = \mathbb{E}\left[\sum_{0<s\leq t}(\Delta N_s)^2\right] = \mathbb{E}\left[\sum_{0<s\leq t}\Delta N_s\right] = \mathbb{E}\left[N_t\right],$$

where the third equality follows because (i) $\lambda\int_0^t R_s \mathrm{d}s$ is continuous in t and thus the jumps of Z coincide with the jumps of N, (ii) the fourth equality follows because $\mathbb{P}(\Delta N_s \in \{0,1\} \forall s \geq 0) = 1$ since multiple individuals cannot die at the same time with strictly positive probability because they are independent and their survival times are continuous random variables and (iii) the last equality follows because $t \mapsto N_t$ is a step function with jump sizes of $+1$. Since under H_0, $\mathbb{E}[N_t] = \lambda\mathbb{E}[\int_0^t R_s \mathrm{d}s]$ as under H_0, $\mathbb{E}[Z_t] = 0$, we thus conclude that under H_0, $\mathrm{Var}(Z_t) = \lambda\mathbb{E}[\int_0^t R_s \mathrm{d}s]$ for any $t \geq 0$.

Answer to 4

(a) Let x_1 be the amount of time an individual has spent in jail and $x_2 = 1$ respectively $x_2 = 0$ if an individual has a high respectively low income. Then the hazard function of an individual with covariates x_1 and x_2 is given by

$$\mu(t) = \mu_0(t)\mathrm{e}^{\beta_1 x_1 + \beta_2 x_2},$$

where $\mu_0(t)$ is the baseline hazard function and β_1, β_2 are the regression coefficients corresponding to the covariate x_1 and x_2 respectively.

(b) We label the persons in the table from left to right by 1 to 5. Denote the multiplier of individual i by

$$\psi_i = \exp(\beta_1 x_1^{(i)} + \beta_2 x_2^{(i)}),$$

where $x_1^{(i)}$ and $x_2^{(i)}$ are the values of the covariates x_1 and x_2 of individual i. Using the data and the definitions of x_1 and x_2 we have

$$\psi_1 = \mathrm{e}^{5\beta_1}, \quad \psi_2 = \mathrm{e}^{2\beta_1}, \quad \psi_3 = \mathrm{e}^{9\beta_1+\beta_2}, \quad \psi_4 = \psi_5 = \mathrm{e}^{10\beta_1+\beta_2}.$$

Since there are no ties with the genuine failure times, we can use the following formula for the partial likelihood that one can find in the notes:

$$L_P(\beta_1, \beta_2) = \prod_{j=1}^{k} \frac{\psi_{\mathtt{I}_j}}{\sum_{m \in R_j} \psi_m},$$

where k is the number of distinct genuine failure times, $\mathtt{I}_j$ denotes the individual who failed at the jth ordered, genuine failure time which is denoted by τ_j and R_j consists of the set of individuals who are at risk of failing just before time τ_j. The values of τ_j, $\mathtt{I}_j$ and R_j are given below:

j	τ_j	$\mathtt{I}_j$	R_j
1	1	1	$\{1,2,3,4,5\}$
2	6	3	$\{3,4,5\}$
3	15	5	$\{5\}$

Therefore,

$$\begin{aligned} L_P(\beta_1, \beta_2) =& \frac{\psi_1}{\psi_1+\psi_2+\psi_3+\psi_4+\psi_5} \times \frac{\psi_3}{\psi_3+\psi_4+\psi_5} \times \frac{\psi_5}{\psi_5} \\ =& \frac{\mathrm{e}^{5\beta_1}\mathrm{e}^{9\beta_1+\beta_2}}{(\mathrm{e}^{5\beta_1} + \mathrm{e}^{2\beta_1} + \mathrm{e}^{9\beta_1+\beta_2} + 2\mathrm{e}^{10\beta_1+\beta_2})(\mathrm{e}^{9\beta_1+\beta_2} + 2\mathrm{e}^{10\beta_1+\beta_2})}. \end{aligned}$$

[Mark distribution: 3 for correct definition of L_P with explanation of notation, 3 for correct table, 1 for correct final answer.]

(c) A left-truncated survival time means that the individual was not observed from time 0 onwards but from a later time onwards. In the context of this study time 0 is the time when the individual was released from jail, so a survival time left-truncated at time t means that the corresponding individual started being observed t months after he/she was released from prison.

(d) As β_2 is the regression coefficient corresponding to the income covariate, we want to test $H_0 : \beta_2 = 0$ versus $H_A : \beta_2 \neq 0$. With $L_P(\beta_1, \beta_2)$ the partial likelihood, the test statistic in the likelihood ratio test takes the form,

$$\Lambda = \frac{\max_{\beta_1} L_P(\beta_1, 0)}{\max_{\beta_1,\beta_2} L_P(\beta_1, \beta_2)}.$$

If $-2\log\Lambda > \lambda_\alpha$ where $\lambda_\alpha > 0$ is such that $\mathbb{P}(\chi_1^2 = \lambda_\alpha) = \alpha$ (with χ_ν^2 denoting a random variable that is chi-squared distributed with ν degrees of freedom), then we reject H_0 at significance level $\alpha \in (0, 1)$. Otherwise, we do not reject H_0. If we reject H_0 at significance level α, then this suggests that, at this particular significance level, there is evidence that $\beta_2 \neq 0$, which means that there is evidence that the income has an effect on the recidivism rate. On the other hand, if we do not reject the H_0 at significance level α, then there is no significant (at level α) evidence that the income has an influence on the recidivism rate.

[Mark distribution: with reference to the question statement, 3 for (i) and (iii), 3 for (ii).]

Answer to 5

(a) Let T be the "aggregate/collective" survival time of the population and let $\alpha(t) = 0.01 \cdot t\mathrm{e}^{-bt}$ so that $\mu(t|Z) = \alpha(t)Z$. By definition of the individual hazard function $\mu(t|Z)$, the individual survival function (given the frailty Z) is given by

$$\mathbb{P}(T > t|Z) = \mathrm{e}^{-\int_0^t \mu(s|Z)\mathrm{d}s} = \mathrm{e}^{-\int_0^t \alpha(s)\mathrm{d}sZ} = \mathrm{e}^{-A(t)Z} \tag{1}$$

where $A(t) := \int_0^t \alpha(s)\mathrm{d}s$ is given by

$$A(t) = \begin{cases} 0.01 \int_0^t s\mathrm{e}^{-bs}\mathrm{d}s = 0.01 \left(-\frac{s}{b}\mathrm{e}^{-bs} - \frac{1}{b^2}\mathrm{e}^{-bs}\right)_{s=0}^t = \frac{0.01}{b^2}\left(1 - bt\mathrm{e}^{-bt} - \mathrm{e}^{-bt}\right) & \text{if } b > 0, \\ 0.01 \int_0^t s\mathrm{d}s = 0.01 \left(\frac{1}{2}t^2\right)_{s=0}^t = 0.005 \cdot t^2 & \text{if } b = 0. \end{cases} \tag{2}$$

The Laplace transform $\phi(\lambda)$ of Z is given by

$$\begin{aligned} \phi(\lambda) = \mathbb{E}[\mathrm{e}^{-\lambda Z}] &= \mathrm{e}^{-\lambda \cdot 1}\mathbb{P}(Z = 1) + \int_0^\infty \mathrm{e}^{-\lambda z} f_Z(z)\mathrm{d}z \\ &= p\mathrm{e}^{-\lambda} + (1 - p)\int_0^\infty \mathrm{e}^{-(\lambda+1)z}\mathrm{d}z \\ &= p\mathrm{e}^{-\lambda} + (1 - p)\frac{1}{\lambda + 1}. \end{aligned} \tag{3}$$

For later reference, note that

$$-\frac{\phi'(\lambda)}{\phi(\lambda)} = \frac{p\mathrm{e}^{-\lambda} + (1 - p)\frac{1}{(\lambda+1)^2}}{p\mathrm{e}^{-\lambda} + (1 - p)\frac{1}{\lambda+1}}. \tag{4}$$

Hence by (1) and (3), the survival function of the population is given by

$$\mathbb{P}(T > t) = \mathbb{E}[\mathbb{P}(T > t|Z)] = \phi(A(t)) = p\mathrm{e}^{-A(t)} + (1 - p)\frac{1}{A(t) + 1}. \tag{5}$$

The connection between the population survival function and the population hazard function denoted by $\mu(t)$ is $\mathbb{P}(T > t) = \mathrm{e}^{-\int_0^t \mu(s)\mathrm{d}s}$. Using this, (5) and (4), we then deduce that the population hazard function $\mu(t)$ is given by

$$\mu(t) = -\frac{\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(T > t)}{\mathbb{P}(T > t)} = -A'(t)\frac{\phi'(A(t))}{\phi(A(t))} = -\alpha(t)\frac{\phi'(A(t))}{\phi(A(t))} = 0.01 \cdot t\mathrm{e}^{-bt}\frac{p\mathrm{e}^{-A(t)} + (1-p)\frac{1}{(A(t)+1)^2}}{p\mathrm{e}^{-A(t)} + (1-p)\frac{1}{A(t)+1}}$$

with $A(t)$ given by (2).

[Mark distribution: 3 for correct general formula with explanation of notation, 2 for correct $\phi(\lambda)$, 2 for correct $A(t)$, 1 for correct final answer.]

(b) The explanation of the research team corresponds to the individual hazard function $\mu(t|Z) = 0.01 \cdot t\mathrm{e}^{-bt}Z$ having the same shape as the population hazard function. Whereas $b = 0$ leads to $\mu(t|Z)$ being increasing over time, $b = 0.05$ leads to a peak of $\mu(t|Z)$ at $t = 20$ (since $t\mathrm{e}^{-bt}$ attains its maximum at $t = 1/b$). If $p = 1$, then individual hazard function and the population hazard function coincide whereas $p = 0.75$ the shape of the population hazard function would be somewhat different from the one of the individual hazard function and the point where the maximum is reached will shift. So the choice $p = 1$ and $b = 0.05$ aligns best with the explanation of the research team.
Note: the choice $p = 0.75$ and $b = 0.05$ is also in quite close alignment with the explanation of the research team and students can get full marks for choosing this set as well (provided their justification is sound).

(c) (i) Due to individuals having different frailty levels (e.g. because of different genes), some individuals are at any time more likely to develop the disease than others. The ones who are more likely to get the disease will typically get the disease at a relatively early age meaning that at a later age the part of the population that is still at risk of developing the disease will consist primarily of less frail individuals who are less likely to obtain the disease and who will therefore drag down the population hazard rate at those later ages. This is the frailty effect which can explain the population hazard rate being decreasing after a certain age. In combination with the disease being less likely to occur at very young ages, the frailty effect can cause a peak of the population hazard rate at e.g. age 20.

(ii) Whereas $b = 0.05$ leads to an individual hazard rate peaking at age 20, the $b = 0$ case gives an increasing individual hazard function which goes against the explanation of the research team. Further, $p = 1$ results in a proportional frailty model where everyone has the same frailty level so the frailty effect cannot take place then. Hence $p = 0.75$ and $b = 0$ is the right answer.

[Mark distribution: 3 for (i), 3 for (ii); though a very good answer in one part showing knowledge/understanding regarding the other part can lead to a higher mark than 3 for a single part if e.g. no answer is given for the other part.]

Contingencies 2, Exam May-June 2018-19 SOLUTIONS

1 Context

The examination follows on from the approach adopted in previous years and is similar to that used for Contingencies 1. A formula sheet is attached to the examination paper which is a subset of relevant information in the Orange Book. It is a shortened version of the sheet for Contingencies 1 with some added joint life functions described.

The exam will be sat in the computer cluster with internet access restricted to the manchester.ac.uk domain only and access to no software other than R, R Studio and Notepad. Supervision, including the presence of IT staff, will be arranged in the same way as in previous years.

As for the questions, they are set consistently with the material covered in the course notes. Questions are also set consistently with the tutorials.

- Question 1 is straightforward book work although it does get progressively harder.
- Question 2 is similar to the questions from the exercise sheet providing a test on the use of the R functions and an appreciation of the short cut in the notes for part b). The use of a same sex couple means that the student reduces the time spent changing the mortality table.
- Question 3 is a simple proof. We don't cover proof work that often in the exams.
- Question 4 is similar to exercise sheet questions and is, for parts a) and b) largely book work. The final part is a bit different as, given constant forces of transition, it is an EPV that can be evaluated.
- Question 5 is a basic evaluation of the probabilities in a multiple decrement table followed by an example where the two decrements do not compete within a one year period.
- Question 6 is a pensions question. Students generally find these to be a bit harder, albeit that the main part is a fairly standard case. More testing are parts b) and c) which requires an onderstanding of how the salary scale works and also the final averaging factor. The question addresses a form of pension benefit change seen in practice.

- Question 7 is different to the questions set in this area in previous year. It requires an appreciation of what the data shows and then asks for an adjustment to be made which is one of the harder aspects of the work on unit linked policies. The question also gives opportunity to do a basic zeroisation calculation.
- Question 8 is book work in the first part and an opportunity to apply the knowledge the students have to a current phenomena.

The material covered in Contingencies 2 is based primarily on objectives (vi) to (x) (but with reference back to (i) to (v)) of the CT5 course. In terms of the questions they are set based on:

- Question 1 - Based mainly on objective (vi), part 1.
- Question 2 - Based mainly on objective (vi), part 1.
- Question 3 - Based mainly on objective (vi), part 1.
- Question 4 - Based mainly on objective (vii), part 3.
- Question 5 - Based mainly on objective (viii), part 2.
- Question 6 - Based mainly on objective (viii), part 6.
- Question 7 - Based mainly on objective (ix), parts 3 and 5.
- Question 8 - Based mainly on objective (x), part 3.

2 Intended Learning Outcomes

1. Develop formula for and calculate expected present values (using either the standard R functions used in the course, by hand or using simple integration techniques) for assurances and annuities for joint life policies, last survivor policies and also policies where the order in which death might occur is a factor. Use these to calculate premiums and reserves for policies.

2. Write down integrals for the expected present value of benefits subject to competing risks (e.g. retirement and death) defined by a Markov model.

3. Recognise the difference between the forces of transition, dependent rates of transition and independent rates of transition and be able to calculate one from another.

4. Use the forces and rates of transition described in 3. correctly in the context of both Multi State Modelling and the development and application of a Multiple Decrement Table to calculate expected present values.

5. Develop formula to calculate the expected present value of defined benefit pension schemes and use the standard functions developed in R to calculate their value.

6. Develop formula for and calculate the profit vector, profit margin and Internal Rate of Return for both conventional and Unit Linked policies.

7. Describe why insurers zeroise their reserves and demonstrate how to do this.

8. Describe the factors affecting mortality and also the different forms of mortality selection.

9. Write down the formula and calculate values for a number of single figure indices developed during the course which are used to compare the mortality experience between different populations.

3 Solutions

1. **QUESTION 1**

   Using the following extracts from two simplified mortality tables, one for life x and one for life y, calculate the probabilities below (in all cases assume the first life is life x and the second is life y):

age x	l_x	age y	l_y
30	10,000	20	10,000
31	9,900	21	9,950
32	9,800	22	9,900
33	9,700	23	9,850

   (a) $p_{32:22}$

   (b) ${}_2p_{30:20}$

   (c) ${}_{2|}q_{30:20}$

   (d) ${}_{2|}q_{\overline{30:20}}$

   [Total 12 marks]

   Distribution of marks : 2, 2, 4, 4, total marks = 12

   ILO covered

   ILO 1: Develop formula for and calculate expected present values (using either the standard R functions used in the course, by hand or using simple integration techniques) for assurances and annuities for joint life policies, last survivor policies and also policies where the order in which death might occur is a factor. Use these to calculate premiums and reserves for policies.

SOLUTION Q1

(a) $p_{32:32} = p_{32} \cdot p_{32} = (9,700/9,800) \times (9,850/9900) = 0.9848$

(b) ${}_2p_{30:30} = (9,800/10,000) \times (9,900/10,000) = 0.9702$

(c) ${}_{2|}q_{30:30} = {}_2p_{30:30} \times (1 - p_{32:32}) = 0.9702 \times (1 - 0.9848) = 0.01475$

(d) ${}_{2|}q_{\overline{30:30}} = {}_{2|}q_{30} + {}_{2|}q_{30} - {}_{2|}q_{30:30} = 0.01 + 0.05 - 0.01475 = 0.00025$

Note the terms for ${}_{2|}q_{30}$ in part d) above are for each of the two lives and equal $(9,800 - 9,700)/10,000$ and $(9,900 - 9,850)/10,000$ respectively.

2. **QUESTION 2** For two lives both aged 40 exact and assuming that both lives are statistically independent, calculate the expected present value of the following assurances, assuming a rate of interest of 3% p.a. and mortality of PFA90C20 for both lives:

   (a) $A_{40:40}$

   (b) $A^1_{40:40}$

   (c) $A_{\overline{40:40}}$

   Distribution of marks : 3, 2, 4, total marks = 9

   ILO covered

   ILO 1: Develop formula for and calculate expected present values (using either the standard R functions used in the course, by hand or using simple integration techniques) for assurances and annuities for joint life policies, last survivor policies and also policies where the order in which death might occur is a factor. Use these to calculate premiums and reserves for policies.

SOLUTION Q2

```
 > useLifeTable("PA92C20")
>
> i<-0.03
> x<-40
> y<-40
>
> ## part a)
>
> A4040<-getEPVJointWholeLifeAss(x,"F",y,"F",i)
> A4040
[1] 0.3029919
>
>
> ## part b) answer is half A4040.
>
> ## part c) Need A40
>
> useLifeTable("PFA92C20")
>
> A40<- getEPVWholeLifeAss(x,i)
> A40
[1] 0.2580503
>
>
> ## answer is 2xA40 - A4040
```

(a) This is: $A_{40:40} = 0.30299$

(b) This is: $A^{1}_{40:40} = 0.5 \times 0.30299 = 0.15150$

(c) This is: $A_{\overline{40:40}} = 2 \times 0.25805 - 0.30299 = 0.2131$

3. **QUESTION 3**

An insurance company issues a policy to two lives x and y which pays a sum assured of £10,000 immediately on the first death. Premiums are payable continuously until the first death occurs.

Assume a mortality model for life x which has $\mu_x = 0.01$ for all $x \geq 0$ and for life y has $\mu_y = 0.012$ for all $y \geq 0$.

(a) Show that the premium for this policy is independent of the force of interest δ.

(b) Calculate the premium.

(c) If the policy was a whole life last survivor policy paying the same sum assured immediately on death, with premiums also payable continuously until the second death, would the amount of premium be dependent on the value of δ? Give an explanation for your answer.

Distribution of marks : 3, 1, 4, total marks = 8

ILO'S covered

ILO 1: Develop formula for and calculate expected present values (using either the standard R functions used in the course, by hand or using simple integration techniques) for assurances and annuities for joint life policies, last survivor policies and also policies where the order in which death might occur is a factor. Use these to calculate premiums and reserves for policies.

SOLUTION Q3

(a)

$$\begin{aligned}
\text{Premium} &= \frac{10{,}000\bar{A}_{xy}}{\bar{a}_{xy}} \\
&= \frac{10{,}000\int_0^\infty \ e^{-\delta t} \ {}_tp_{xy} \ (\mu_{x+t}+\mu_{y+t})\cdot dt}{\int_0^\infty \ e^{-\delta t} \ {}_tp_{xy}\cdot dt} \\
&= 10{,}000(\mu_{x+t}+\mu_{y+t})\frac{\int_0^\infty \ e^{-\delta t} \ {}_tp_{xy} \ \cdot dt}{\int_0^\infty \ e^{-\delta t} \ {}_tp_{xy}\cdot dt} \\
&= 10{,}000(\mu_{x+t}+\mu_{y+t})
\end{aligned}$$

as $\mu_{x+t}+\mu_{y+t}$ is constant and hence the premium is independent of δ.

(b) Premium is $10{,}000(0.01+0.012) = 220$.

(c) It is dependent on δ.

The premium is calculated as :

$$\frac{\bar{A}_x+\bar{A}_y-\bar{A}_{xy}}{\bar{a}_x+\bar{a}_y-\bar{a}_{xy}}$$

For both numerator and denominator it is not possible to factor out an expression for the force of mortality across the **whole** of the term - for example, in the numerator, the first term has μ_{x+t} as a factor, the second term has μ_{y+t} as a factor and the third term has $\mu_{x+t}+\mu_{y+t}$ as a factor.

4. **QUESTION 4**

Consider a 3-state sickness model with states H (healthy), S (sick) and D (dead) and the following forces of transition for a life aged x

- μ_x^{HS} on transition from healthy to sick,
- μ_x^{HD} on transition from healthy to dead,
- μ_x^{SH} on transition from sick to healthy,
- μ_x^{SD} on transition from sick to dead.

(a) Write down an integral expression for the expected present value (EPV) of a lump sum sickness benefit of £20,000 payable immediately on becoming ill for the first time for a healthy life aged 40 exact.

(b) Write down an integral expression for the expected present value (EPV) of a sickness benefit of £10,000 payable immediately on becoming ill for the second or any subsequent time for a healthy life aged 40 exact.

(c) Assume the above transition intensities take the following constant values, for all $x \geq 0$:

- $\mu_x^{HS} = 0.10$
- $\mu_x^{HD} = 0.01$
- $\mu_x^{SH} = 0.08$
- $\mu_x^{SD} = 0.02$

Calculate the EPV for a sickness benefit of £2,000 p.a. for a life aged 30 exact *who is currently in the sick state* where the benefit is paid until the life either recovers or dies. Assume a force of interest of 0.03.

Distribution of marks : 3, 5,6, total marks = 14

ILO covered

ILO 2: Write down integrals for the expected present value of benefits subject to competing risks (e.g. retirement and death) defined by a Markov model.

ILO 3: Recognise the difference between the forces of transition, dependent rates of transition and independent rates of transition and be able to calculate one from another.

SOLUTION Q4

(a) The integral is $20{,}000 \int_0^\infty \; e^{-\delta t} \; {}_tP_{40}^{\overline{HH}} \cdot \mu_{40+t} \cdot dt$.

(b) The integral is $10{,}000 \int_0^\infty \; e^{-\delta t} \; \left({}_tP_{40}^{HH} - {}_tP_{40}^{\overline{HH}}\right) \cdot \mu_{40+t} \cdot dt$.

(c) The EPV is:

$$
\begin{aligned}
EPV &= 2{,}000 \int_0^\infty \; e^{-\delta t} \; {}_tP_{30}^{\overline{SS}} \cdot dt \\
&= 2{,}000 \int_0^\infty \; e^{-0.03t} \; e^{-\int_0^t (\mu_{x+s}^{SH} + (\mu_{x+s}^{SD}) \cdot ds} \cdot dt \\
&= 2{,}000 \int_0^\infty \; e^{-0.03t} \; e^{-\int_0^t (0.08+0.02) \cdot ds} \cdot dt \\
&= 2{,}000 \int_0^\infty \; e^{-0.03t} \; e^{-0.10t} \cdot dt \\
&= 2{,}000 \int_0^\infty \; e^{-0.013t} \cdot dt \\
&= 2{,}000 \big[e^{-0.013t}\big]_0^\infty / -0.13 \\
&= 2{,}000/0.13 = 15{,}385.
\end{aligned}
$$

5. **QUESTION 5**

Consider a 3-state multiple decrement model with states a (active), r (retired) and d (dead).

(a) Given an independent probability of retiring when aged 60 of 0.1 and a constant force of mortality of 0.01 for all ages between 60 exact and 65 exact, calculate $(aq)^d_{60}$ and $(aq)^r_{60}$

(b) If retirement can only occur at the end of a year what is the probability of a life aged 60 exact being active at age 61 exact?

Distribution of marks : 8, 4, total marks = 12

ILO covered

ILO 4: Use the forces and rates of transition described in 3. correctly in the context of both Multi State Modelling and the development and application of a Multiple Decrement Table to calculate expected present values.

SOLUTION Q5

(a) Using the following:

$$q^r_{60} = 0.1, \text{hence } \mu^r_{60} = -ln(1 - q^r_{60}) = 0.10536$$

and

$$\begin{aligned}(aq)^d_{60} &= \frac{\mu^d_{60}}{\mu^r_{60} + \mu^d_{60}} \cdot (1 - e^{-(\mu^r_{60} + \mu^d_{60})}) \\ &= (0.01/0.11536) \cdot (1 - e^{0.11536}) = 0.009445\end{aligned}$$

and

$$\begin{aligned}(aq)^d_{60} &= \frac{\mu^r_{60}}{\mu^r_{60} + \mu^d_{60}} \cdot (1 - e^{-(\mu^r_{60} + \mu^d_{60})}) \\ &= (0.10536/0.11536) \cdot (1 - e^{0.11536}) = 0.09951\end{aligned}$$

(b) This is $(1 - q^d_{60}) \cdot (1 - q^r_{60}) = e^{-0.01} \times 0.9 = 0.89104$

6. **QUESTION 6**

A pension scheme provides a pension on normal health retirement, calculated as follows:

Annual amount of pension payable for life from norml health retirement at age $x = \dfrac{N}{80} \times FPS_x$

where

- FPS_x = the average salary earned over the three years ending at age x and
- N is the total period of time the life has been a member of the pension scheme.

Now, consider a person aged 50 exact who joined the pension scheme 20 years ago. You are told that the person has earned a salary of £25,000 in the previous year.

(a) Using the data and functions provided in R calculate the expected present value (EPV) of the pension assuming the member retires between the age of 63 exact and age 65 exact.

(b) The pension scheme is now changed so that there is a maximum value placed on FPS_x of £30,000. Calculate the EPV of the pension payable , assuming the member retires between the age of 63 exact and age 65 exact, after this change.

(c) Instead of FPS_x having a maximum value of £30,000, S_x, the salary earned between ages x and $x + 1$, has this same maximum amount applied. Does this change:

   i. increase the EPV calculated in part b), or
   ii. decrease the EPV calculated in part b), or
   iii. leave the EPV calculated in part b) unchanged?

Distribution of marks : 14, 4, 4, total marks = 22

ILO covered

ILO 5: Develop formula to calculate the expected present value of defined benefit pension schemes and use the standard functions developed in R to calculate their value.

SOLUTION Q6

See R code:

```
 useLifeTable("Pension")
>
> i<-0.04
>
> ## part a)
>
> #Service vector
>
> servvec<-c(33.5,34.5,35)
> servvec
[1] 33.5 34.5 35.0
>
> # salary vector
>
> z<-getAveSalScale("allHalf")

> s<-getSalScale("all")

> salvec<-25000*z[63:65]/s[49]
> salvec
[1] 30189.53 30649.52 30884.36
>
> # epv of pounds 1 vector
>
> cra<-getCommCr("all",i)

> D<-getCommD("all",i)
> D[50]
[1] 1795.915
>
> # overall EPV
>
> EPV<-servvec*salvec*cra[63:65]/D[50]
> EPV
[1]  18432.98  16044.74 176683.10
> Ansa<-sum(EPV)
>
> # Answer to part a)
```

```
>
> Ansa
[1] 211160.8


>
> ## part b)
>
> # see what FPS is
>
> salvec
[1] 30189.53 30649.52 30884.36
>
> # all values above pounds 30k
>
> ansb<-sum(30000*servvec*cra[63:65]/D[50])
>
> # Answer to part b)
>
> ansb
[1] 205645.8
>
> ## part c)
>
> # see all salary figures from S60 to S64 as these impact z63.5, z64.5 and z65
>
> svec<-25000*s[60:64]/s[49]
> svec
[1] 29517.77 29949.62 30425.76 30868.67 31358.65
>
> # some are less than pounds 30k so 3 year average salary is less than pounds 30k and EPV is less than for b)
```

7. **QUESTION 7**

An insurer writes a 5 year unit linked endowment policy for a life aged 35. Death benefits are payable at the end of the year of death and the maturity value on completion of the 5 year term. Level premiums are paid yearly in advance. Benefits are paid at the end of the year.

You are given the following tables of data:

Unit Fund

year	Premium paid	Premium allocated	bid offer spread	net premium allocated to policyholder	fund before management charge	management charge	Unit Fund at year end
1	10,000	8,000	320	7,680	8,064	81	7,983
2	10,000	10,500	420	10,080	18,966	190	18,776
3	10,000	10,500	420	10,080	30,299	303	29,996
4	10,000	10,500	420	10,080	42,080	421	41,659
5	10,000	10,500	420	10,080	54,326	543	53,783

Non-Unit Fund

year	Non allocated premium	bid/offer spread	Expenses	interest on fund at 4%	management charge	Profit
1	2,000	320	3070	(30)	81	(699)
2	(500)	420	120	(8)	190	(18)
3	(500)	420	120	(8)	303	95
4	(500)	420	120	(8)	421	213
5	(500)	420	120	(8)	543	335

In addition assume :

- $(ap)_x = 0.99$ for all x
- Interest on reserves = 3%p.a.

(a) State the allocation percentage and the bid offer spread (as a percentage) for this policy for all years.

(b) What is the assumed investment return on the Unit Fund each year implied by this data?

(c) What is the amount payable on death implied by this data?

(d) Zeroise the profit vector for this policy

(e) If the insurance company decides to introduce a penalty to the policyholder on surrender equal to 10% of the value of the Unit Fund, what is the new Profit Vector. Assume the dependent probability of surrender, $(aq)_x^s = 0.05$ for all ages x.

Distribution of marks :2, 3, 2, 6, 4, total marks = 17

ILO's covered

ILO 6: Develop formula for and calculate the profit vector, profit margin and Internal Rate of Return for both conventional and Unit Linked policies.

ILO 7: Describe why insurers zeroise their reserves and demonstrate how to do this.

SOLUTION Q7

(a) 80% in year one and then 105% thereafter. The bid offer spread is 4%

(b) 5% for all years.

(c) The death benefit is the value of the Unit Fund at the end of the year of death.

(d) We use the equuation below:

$$(NUCF)_t \ + \ {}_{t-1}V \ (1+i) \ - \ p_{x+t-1} \ {}_tV \ = \ (PRO)_t$$

As the last negative cash flow is in year 2 we can say that ${}_2V = 0$.

Hence

$$\text{As } (NUCF)_2 \; - \; p_{36} \; {}_2V < 0, (PRO)_2 = 0$$

$$\begin{aligned} \text{Calculate } {}_1V : -18 + {}_1V \times 1.03 - p_{33} \times 0 &= 0 \\ {}_1V &= 18/1.03 \\ &= 17.48 \end{aligned}$$

$$\text{As } {}_0V = 0, \text{by definition}$$

$$\begin{aligned} \text{Calculate } (PRO)_1 : -699 + 0 \times 1.03 - p_{30} \times 17.48 &= (PRO)_1 \\ (PRO)_1 &= -699 - 0.99 \times 17.48 \\ &= -716.30 \end{aligned}$$

Hence the zeroised Profit Vector is:

$$(-716, 0, 95, 213, 335)$$

(e) the profit from surrenders is calculated as Unit Fund $\times 0.1 \times 0.05$.

Hence the vector of profits on surrender is (40, 94, 150, 208, 0). The last term is zero because the policy matures after 5 years.

Applying these to the non-zeroised vector (over reserving will otherwise occur) gives the profit vector :

(-659, 76, 245, 421, 335)

8. **QUESTION 8**

   (a) List four factors that affect mortality.

   (b) Mortality for a population has, for 20 years, consistently shown improvements. Recently these improvements have stopped. Referring to the factors affecting mortality, give four reasons why this might have happened.

   Distribution of marks : 2, 4, total marks = 6

   ILO covered

   ILO 8:Describe the factors affecting mortality and also the different forms of mortality selection.

   SOLUTION Q8

   (a) 4 from: nutrition, occupation, lifestyle, climate, education, housing or any other reasonable answer.

   (b) any reasonable comment such as : More sedentary lifestyles, worse diet, climate change/environment may cause more pollution, poor education regarding lifestyle issues, poorer healthcare etc.

Solutions for MATH39542 Risk Theory exam 2018/2019

Question justification

Most questions are similar to exercise sheets questions. The degree of similarity varies somewhat, however no exam question is identical to an exercise sheet question. Questions 1(d), 2(d) and 4(b) are unseen variations on their theme. Questions 1(b), 2(e) and 4(c) should be considered new/unseen. Questions 3(a), 5(b) and 5(c) are bookwork.

IFoA syllabus coverage

This course covers the IFoA syllabus objectives listed in below table.

CT	unit	topic	exam question(s)
6	4	Ruin Theory	1 & 2
6	5	Bayesian Statistics	4
6	6	Credibility Theory	5

ILOs

The ILOs for this course are as follows. Be able to

1. compare and contrast the Bayesian approach and the frequentist approach to Statistics.
2. describe the Bayesian approach to Decision Theory and relate the concepts Bayes risk, Bayes rule, posterior risk and risk function.
3. determine the optimal action or decision function for a given decision problem using the Bayesian approach.
4. evaluate Bayes estimate for the unknown parameter in the context of Bayesian statistical inference.
5. discuss the assumptions and goals of Credibility Theory, and compute the types of credibility estimates seen in the course.
6. describe the premium principles and their desirable properties seen in the course, and (dis)prove whether a premium principle satisfies one of these desirable properties for examples of similar difficulty as seen in the course.
7. discuss the elements making up Cramer-Lundberg model and the Net Profit Condition.
8. evaluate the probability of ruin in a Cramer-Lundberg capital process for examples of similar difficulty as seen in the course.
9. discuss the key idea in the proof of Lundberg's inequality and apply it to a Cramer-Lundberg capital process.
10. explain how reinsurance can be incorporated in the Cramer-Lundberg model and determine the optimal level of reinsurance for examples of similar difficulty as seen in the course.

Coverage of these ILOs by the exam questions is as follows (low/medium/high):

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
q1(a),(b),(c)							m			
q1(d)							h		h	
q2(a)							m			
q2(b),(c),(d)								h		
q2(e)										h
q3(a)						l				
q3(b)						m				
q4(a)				m						
q4(b),(c)		h	h	h						
q5		m			m					

Note: ILO 1 was covered in the first piece of coursework for this course.

Solutions

Question 1

(a) For the trajectories, see Figure 1. As the trajectory for $(U_t)_{t\geq 0}$ shows, ruin happens after 9 months.

(b) The probability of ruin during the first year decreases, since the capital process with this change in place dominates $(U_t)_{t\geq 0}$.

(c) The NPC entails that $c > \lambda\mathbb{E}[X_1]$. Using that we have seen in the exercises that $\mathbb{E}[S_t] = \lambda t\mathbb{E}[X_1]$ we indeed get for any $t > 0$ that

$$\mathbb{E}[U_t] = u + ct - \mathbb{E}[S_t] = u + ct - \lambda t\mathbb{E}[X_1] = u + t\big(c - \lambda\mathbb{E}[X_1]\big) > u.$$

(d) We are asked to compute $\mathbb{P}(U_{J_1} < 0)$, where $J_1 \sim \mathrm{Exp}(\lambda)$ denotes the arrival time of the first claim. Conditioning on J_1:

$$\begin{aligned}\mathbb{P}(U_{J_1} < 0) = \mathbb{P}(u + cJ_1 - X_1 < 0) &= \mathbb{E}[\mathbb{P}(u + cJ_1 - X_1 < 0 \mid J_1)] \\ &= \mathbb{E}[1 - F_{X_1}(u + cJ_1)] = \int_0^\infty \big(1 - F_{X_1}(u + cs)\big)\lambda e^{-\lambda s}\,\mathrm{d}s.\end{aligned}$$

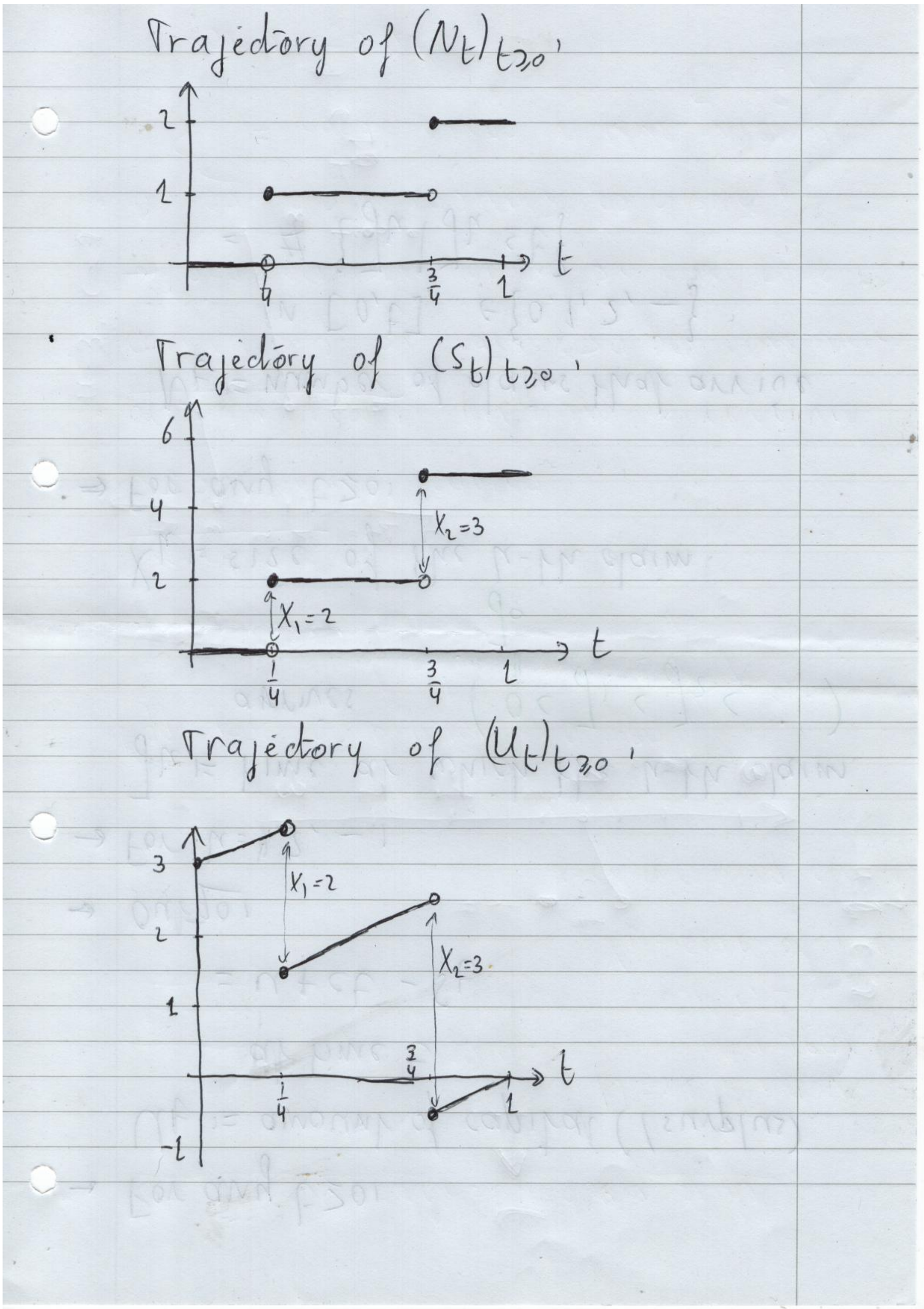


Figure 1: Trajectories for question 1(a)

Question 2

(a) The NPC entails that $c > \lambda\mathbb{E}[X_1]$. We can compute

$$\mathbb{E}[X_1] = \int_0^\infty x f(x)\,\mathrm{d}x = \frac{1}{8}\frac{1}{2}\int_0^\infty 2xe^{-2x}\,\mathrm{d}x + \frac{15}{4}\frac{1}{4}\int_0^\infty 4xe^{-4x}\,\mathrm{d}x$$
$$= \frac{1}{8}\frac{1}{2}\frac{1}{2} + \frac{15}{4}\frac{1}{4}\frac{1}{4} = \frac{17}{64}$$

and hence the NPC reads $1/2 > 1 \cdot 17/64$ and this indeed holds. For the integral we used the mean of Exponetial distributions (also available on attached distribution list), otherwise the integral can also be computed using integration by parts e.g.

(b) We know from the lecture (notes) that the Laplace exponent can be expressed as

$$\xi(\theta) = c\theta - \lambda + \lambda\mathcal{L}_{X_1}(\theta) \quad \text{for } \theta \in \mathbb{R}.$$

Here

$$\mathcal{L}_{X_1}(\theta) = \mathbb{E}[e^{-\theta X_1}] = \frac{1}{8}\int_0^\infty e^{-(\theta+2)x}\,\mathrm{d}x + \frac{15}{4}\int_0^\infty e^{-(\theta+4)x}\,\mathrm{d}x$$
$$= \begin{cases} \infty & \text{if } \theta \le -2 \\ \frac{1}{8(\theta+2)} + \frac{15}{4(\theta+4)} & \text{if } \theta > -2. \end{cases}$$

Plugging in that $c = 1/2$ and $\lambda = 1$ we hence get that

$$\xi(\theta) = \begin{cases} \infty & \text{if } \theta \le -2 \\ \frac{1}{2}\theta - 1 + \frac{1}{8(\theta+2)} + \frac{15}{4(\theta+4)} & \text{if } \theta > -2 \end{cases}$$

and its domain is hence $(-2, \infty)$.

(c) The Lundberg coefficient R is defined as the absolute value of the only strictly negative root of ξ (provided it exists). Hence we need to work out $\xi(\theta) = 0$, which with some algebra leads to

$$\left(\frac{1}{2}\theta - 1\right) 8(\theta+2)4(\theta+4) + 4(\theta+4) + 15 \cdot 8(\theta+2) = 0 \implies$$
$$16\theta^3 + 64\theta^2 + 60\theta = 0 \implies 4\theta\left(4\theta^2 + 16\theta + 15\right) = 0 \implies \theta \in \left\{-\frac{5}{2}, -\frac{3}{2}, 0\right\}.$$

Noting that $-3/2$ is the only strictly negative solution inside the domain of ξ, it follows that $R = 3/2$.

(d) From the lecture (notes) we know that the ruin probability $u \mapsto \psi(u)$ is a non-increasing function, that $\psi(0) = \lambda\mathbb{E}[X_1]/c = 1 \cdot (17/64)/(1/2) = 17/32$ (the mean of

X_1 was already computed in part (a)) and that (Lundberg's inequality) $\psi(u) \le e^{-Ru}$ for all $u \ge 0$.

Now, ψ_1 violates (only) Lundberg's inequality (no matter whether the correct value for R from part (b) or otherwise the suggested substitute is used), ψ_2 violates (only) that the value in $u = 0$ should be $17/32$, and ψ_3 violates (only) the non-increasing property (indeed $\psi_3'(0) = 1/16 > 0$). Hence only ψ_4 remains.

Note: indeed ψ_4 is the correct expression. We know from the lecture (notes) that the Laplace transform of the survival probability satisfies

$$(\mathcal{L}\varphi)(\theta) = \frac{c - \lambda\mathbb{E}[X_1]}{\xi(\theta)} = \frac{15(\theta+2)(\theta+4)}{32\theta(\theta+3/2)(\theta+5/2)} \quad \text{for } \theta > 0$$

(where the expression for ξ from part (b) was reused). It is easy to check that the Laplace transform of $u \mapsto 1 - \psi_4(u)$ equals the above rhs for $\theta > 0$ which confirms that ψ_4 is the correct expression by uniqueness of Laplace transforms. Students have seen these techniques as well, and of course they could use these as well to come to their answer if they so wish.

(e) We know from the lecture (notes) that for $u = 0$ the ruin probability is given by $\lambda\mathbb{E}[X_1]/c$ (prvided that the NPC holds). Considering things as a function of $c \in (0, 3)$, the information about the claim sizes yields that the common mean of the claim sizes can be expressed as

$$\left(1 + 0.2\left(c - \frac{1}{2}\right)\right)\mathbb{E}[X_1] = (0.2c + 0.9)\,\frac{17}{64}.$$

Further, if the interarrival times with a common $\mathrm{Exp}(\lambda) = \mathrm{Exp}(1)$ distribution are mutiplied by a factor h then the resulting times have a common $\mathrm{Exp}(1/h)$ distribution (students may use this as a fact since it was used in the exrecises in a different context, if they don't remember this then they could just work it out) — in particular the resulting process is still a C-L process with claim intensity $1/h$. Hence the information about how the interarrival times depend on c translates to a claim intensity of

$$\left(1 + 0.3\left(\frac{1}{2} - c\right)\right)^{-1} = (1.15 - 0.3c)^{-1}.$$

Putting these together we hence get that the ruin probability is for any $c \in (0, 3)$ given by

$$\frac{0.2c + 0.9}{c(1.15 - 0.3c)}\frac{17}{64} \tag{1}$$

provided that this quantity is in $(0, 1]$, otherwise the NPC does not hold and then the ruin probability is 1. Differentiating (1) wrt c gives

$$\frac{17}{64}\frac{c(1.15 - 0.3c)\cdot 0.2 - (0.2c + 0.9)(1.15 - 0.6c)}{c^2(1.15 - 0.3c)^2} = \frac{17}{64}\frac{0.06c^2 + 0.54c - 1.035}{c^2(1.15 - 0.3c)^2}.$$

The second order polynomial in the numerator has roots

$$\frac{-0.54 - \sqrt{0.54}}{0.12} \approx -10.6 \quad \text{and} \quad \frac{-0.54 + \sqrt{0.54}}{0.12} \approx 1.6$$

and since this polynomial is negative between its roots it follows that (1) attains its minimum on the interval $(0, 3)$ in (approx) 1.6. We still need to check that plugging $c = 1.6$ into (1) does indeed yield a value of less than 1 (otherwise the NPC would not hold in this point — though indirectly it is clear that this can't happen since the NPC also holds for $c = 1/2$, cf. part (a)), and we find a value of approx 0.3.

So we can conclude that the optimal choice for the premium rate is $c = 1.6$, with a corresponding ruin probability of 0.3 (down from approx 0.53 for $c = 1/2$, cf. part (d)).

Question 3

(a) The Esscher premium principle is the mapping

$$\pi(X) = \frac{\mathbb{E}[Xe^{\beta X}]}{\mathbb{E}[e^{\beta X}]} \quad \text{for some } \beta > 0$$

and the Monotonicity property entails that

$$\pi(X) \leq \pi(Y) \quad \text{for all risks } X, Y \text{ so that } X \leq Y.$$

(b) Note: one mark each is given for correctly quoting the principle and properties, two marks for shwoing the first property, three marks for disproving the second property.

The Standard Deviation principle with loading factor $\beta = 1.1$ is the mapping

$$\pi(X) = \mathbb{E}[X] + 1.1\sqrt{\mathrm{Var}(X)}$$

The Scale Invariance property reads as $\pi(cX) = c\pi(X)$ for constants $c > 0$, and indeed

$$\pi(cX) = \mathbb{E}[cX] + 1.1\sqrt{\mathrm{Var}(cX)} = c\mathbb{E}[X] + 1.1\sqrt{c^2\,\mathrm{Var}(X)} = c\pi(X).$$

The No Rip-off property entails that if $X \leq C$ for some constant C, then $\pi(X) \leq C$. Let e.g. X take values 0 and 1 with prob $1/2$ each, and $C = 1$. Then $X \leq C$ holds but

$$\pi(X) = \frac{1}{2} + 1.1\sqrt{\frac{1}{4}} = 1.05 > 1.$$

Question 4

(a) We are asked to compute Bayes estimate for θ under 0-1 loss, which we know from the lecture (notes) is any maximiser of the posterior pmf. Using Bayes Thm to determine the posterior pmf yields (using the attached distribution list as necessary)

$$\begin{aligned} f_{\Theta|X}(\theta|7) &= c f_{X|\Theta}(7|\theta) f_{\Theta}(\theta) \\ &= \begin{cases} c \cdot \binom{10}{7} 0.7^7 0.3^3 \cdot 1/2 = \widetilde{c} \cdot 0.7^4 = 0.7^4/(0.7^4 + 0.3^4) \approx 0.97 & \text{if } \theta = \theta_1 \\ c \cdot \binom{10}{7} 0.3^7 0.7^3 \cdot 1/2 = \widetilde{c} \cdot 0.3^4 = 0.3^4/(0.7^4 + 0.3^4) \approx 0.03 & \text{if } \theta = \theta_2, \end{cases} \end{aligned}$$

where $c, \widetilde{c}$ are constants indepedent of θ and the final equality uses that the probs must sum to 1.

Hence θ_1 has the larger probability and is therefore Bayes estimate for θ under 0-1 loss.

(b) We know from the lecture (notes) that for a general loss function, Bayes estimate is any mimimiser of the posterior risk

$$a \mapsto \mathbb{E}[l(\Theta, a) \,|\, X = 7] \quad \text{for } a \in \{\theta_1, \theta_2\}.$$

Using the posterior distribution from part (a), we get for $a = \theta_1$

$$\mathbb{E}[l(\Theta, a) \,|\, X = 7] = f_{\Theta|X}(\theta_1|7) + c f_{\Theta|X}(\theta_2|7) = 0.97 + 0.03c$$

and for $a = \theta_2$

$$\mathbb{E}[l(\Theta, a) \,|\, X = 7] = c f_{\Theta|X}(\theta_1|7) + f_{\Theta|X}(\theta_2|7) = 0.97c + 0.03.$$

It is readily checked that for any $c > 1$, the former is smaller than the latter.

(c) The quickest way to do this is to realise that the loss function l from part (b) can be related to the 0-1 loss function $l^{0\text{-}1}$ via

$$l(\theta, a) = 1 + (c - 1) l^{0\text{-}1}(\theta, a),$$

and hence the posterior risks are related via

$$\mathbb{E}[l(\Theta, a) \,|\, X = x] = 1 + (c - 1)\mathbb{E}[l^{0\text{-}1}(\Theta, a) \,|\, X = x]$$

from which we can readily conclude that both posterior risks necessarily have the same minimiser a.

Alternatively a student could work out both these posterior risks, in which the pmf of the posterior is now unknown, and appropriately compare these expressions to come to the same conclusion.

Question 5

(a) We have a Bayesian model in which the prior is $\Theta \sim \text{Gamma}(\alpha, \beta)$ and the likelihood is $X_i|\Theta = \theta \sim \text{Exp}(\theta)$ (slight abuse of notation).

We know from the lecture (notes) that the Bayes credibility estimate is given by

$$\widehat{\mu}_B(\vec{x}) = \mathbb{E}[\mu(\Theta) \,|\, \vec{X} = \vec{x}],$$

where μ is the individual premium given by

$$\mu(\theta) = \mathbb{E}[X_i \,|\, \Theta = \theta] = \frac{1}{\theta} \quad \text{for } \theta > 0$$

where we used the $\text{Exp}(\theta)$ distrinbution of the likelihood. Hence

$$\widehat{\mu}_B(\vec{x}) = \mathbb{E}[\Theta^{-1} \,|\, \vec{X} = \vec{x}]. \tag{2}$$

Using Bayes Thm to determine the posterior distribution we get for $\theta > 0$ (using the attached distribution list as necessary)

$$\begin{aligned} f_{\Theta|\vec{X}}(\theta|\vec{x}) &= c f_{\vec{X}|\Theta}(\vec{x}|\theta) f_\Theta(\theta) \\ &= c \cdot \prod_{i=1}^n \theta e^{-x_i\theta} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &= \widetilde{c}\theta^{\alpha+n-1} e^{-(\beta+n\bar{x})\theta}, \end{aligned}$$

where $c, \widetilde{c}$ are constants indepedent of θ and $\bar{x}$ denotes the sample average. We can recognise this as the pdf of a Gamma distribution (where we know we don't need to check multiplicative constants) and can conclude that the posterior has a $\text{Gamma}(\alpha_0, \beta_0)$ distribution for $\alpha_0 := \alpha + n > 1$ and $\beta_0 := \beta + n\bar{x}$.

Finally (2) can now indeed be computed as

$$\begin{aligned} \widehat{\mu}_B(\vec{x}) &= \int_0^\infty \theta^{-1} \frac{{\beta_0}^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0\theta} \, \mathrm{d}\theta \\ &= \frac{{\beta_0}^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_0-1)}{{\beta_0}^{\alpha_0-1}} \int_0^\infty \frac{{\beta_0}^{\alpha_0-1}}{\Gamma(\alpha_0-1)} \theta^{\alpha_0-2} e^{-\beta_0\theta} \, \mathrm{d}\theta \\ &= \frac{\beta_0}{\alpha_0 - 1}, \end{aligned}$$

where the integral on the second line equals 1, and the final equality uses the hint.

(b) We know from the lecture (notes) that both the Bayes and Bühlmann credibility estimate are defined as Bayes estimate for $\mu(\theta)$ under squared error loss, the only difference being that Bühlmann minimises the Bayes risk involved over linear (in $\vec{x}$) decision functions only. However part (a) shows that the minimiser in the class of all decision functions is linear in $\vec{x}$, and hence minimising over only linear decision functions will yield the same result.

(c) One advantage is that the Bühlmann credibility estimate does not require the posterior distribution, in particular it can still be used if only the mean and the variance of the prior is available.