Name: ..... Student card number: .....

Hand in this sheet only.

Rules

- ID required.
- You are not allowed to leave the exam room during the first 30 minutes.
- Scratch paper is handed out. You cannot use your own. It is possible to request additional scratch paper from the invigilator.

  The use of markers is not permitted.
- If you want to go to the toilet, raise your finger to warn a security guard. He or she will give you permission to go and walk with you to the toilet. Toilet visits are not permitted during the first and last half hour of the exam. You may only visit the toilet once.

  It is forbidden to take a telephone or similar electronic devices to the toilet.
- After you have left the examination room, you are not allowed to stay in the corridors / hall immediately outside due to noise. You follow the instructions of the invigilator.

Instructions

- There are open questions and multiple-choice questions.
- Every multiple-choice question has exactly one correct answer. In some cases, other answers may be "almost correct " or "partly correct". In such cases the best answer applies.

  Answer in the appropriate boxes by placing a cross. If you make a mistake, scratch the cross and put a cross in another box.

  Each correctly answered multiple-choice item yields one point.
- Answers to open questions are entered in the boxes (open rectangles)

  First draft your answer. Then fill the box.

  Each correctly answered open item yields two points, unless indicated otherwise.
- Because there are different versions of the exam, the order of the multiple-choice questions does not always correspond with the order of the material as discussed in the lectures.
- It is possible to request a new answer sheet as well as additional scratch paper from the invigilator. Our stock of answer sheets is finite, first come first serve.

  Good luck!

Multiple-choice answers

	A	*B*	*C*	*D*
1.				
2.				
3.				
4.				

	A	*B*	*C*	*D*
5.				
6.				
7.				
8.				

	A	*B*	*C*	*D*
9.				
10.				
11.				
12.				

Open questions—first draft your answer, then fill the box

1. Give four performance standards for MAL algorithms followed by descriptions. (Six were discussed in the lectures.)

Answer. See the introductory slides, and/or the corresponding chapter of "Multi-agent systems" (Y. Shoham and K. Leyton-Brown).

Here are six:

(a) **Auto-compatible**. Approximate Pareto-optimality in self-play.

(b) **Safety**. At least earn the maxmin (security value).

(c) **Targeted optimality**. Best response against a limited class of opponents.

(d) **Efficient targeted learning**. For every $\epsilon > 0$ and $0 < \delta < 1$, there exists an M polynomial in $1/\epsilon$ and $1/\delta$, such that after M steps, with probability $\geq 1 - \delta$, (a), (b) and (c) are achieved within ϵ.

(e) **Rational**. Approximate a best response if the opponents settle on stationary strategies.

(f) **No regret**. At any point, earn no less than any pure strategy would have.

2. (4 points). Two players repeat the following game an indefinite number of times.

	L	R
T	$3, 0$	x, y
B	$0, 1$	$1, 4$

The probability to continue is $0 \leq \delta < 1$. The strategy of the row player is to alternate between T and B (starting with T) as long as col alternates between L and R (starting with L). Row falls back to playing T forever when col does not comply. Analogous for the column player, with fall back strategy playing R forever.

For which values of x, y and δ the strategies just described form a Nash equilibrium in the repeated game? The box should contain (a summary of) a derivation of the answer.

Answer. See the slides on repeated games, and/or the corresponding chapter of "Game Theory: A Multi-Leveled Approach" (H. Peters).

A reason for row to defect would be when it is supposed to play B and play T instead—it would then earn x instead of 1. Row's expected payoff from then on:

$$x + x\delta + x\delta^2 + \ldots = x(1 + \delta + \delta^2 + \ldots) = x\frac{1}{1-\delta}.$$

Row's expected payoff when it complies (starting at the same round):

$$\begin{aligned} 1 + 3\delta + 1\delta^2 + 3\delta^3 + \ldots &= (1 + 3\delta)(1 + \delta^2 + \delta^4 + \delta^6 + \ldots) \\ &= (1 + 3\delta)\frac{1}{1-\delta^2}. \end{aligned}$$

So for row not to deviate, it must be that

$$(1 + 3\delta)\frac{1}{1-\delta^2} > x\frac{1}{1-\delta},$$

which means

$$x < 1 \text{ or } \left(1 \leq x < 2 \text{ and } \frac{1-x}{x-3} < \delta\right).$$

If $x \geq 2$ then x is too large to prevent row from deviating.

A reason for col to defect would be when it is supposed to play L and play R instead—it would then earn y instead of 0. Col's expected payoff from then on would be:

$$y + y\delta + y\delta^2 + \ldots = y\frac{1}{1-\delta}.$$

Col's expected payoff when it complies (starting at the same round):

$$0 + 4\delta + 0\delta^2 + 4\delta^3 = 4\delta(1 + \delta^2 + \delta^4 + \dots) = 4\delta\frac{1}{1-\delta^2}$$

So for row not to deviate, it must be that

$$4\delta\frac{1}{1-\delta^2} > y\frac{1}{1-\delta},$$

which means

$$y < 0 \text{ or } \left(0 \leq y < 2 \text{ and } \frac{y}{4-y} < \delta\right).$$

If $y < 0$ then y is of too little value for col to consider deviating, no matter the value of δ. If $y \geq 2$ then y is of too much value for col to prevent it from deviating, no matter the value of δ.

Summarized:

$$\begin{cases} x < 1 \text{ or } (1 \leq x < 2 \text{ and } \dfrac{1-x}{x-3} < \delta)\ ; \\ y < 0 \text{ or } (0 \leq y < 2 \text{ and } \dfrac{y}{4-y} < \delta)\ . \end{cases}$$

Multiple-choice questions

1. Non-linear Cournot dynamics demonstrates the following.

   (a) MAL does not always converge.

   $\surd$ MAL may become chaotic.

   (c) MAL metrics need not be linear.

   (d) MAL may involve teaching.

> **Explanation.** Cf. slides introduction and Tönu Puu. Chaos in Duopoly Pricing. Chaos, Solitions & Fractions 1(6), pp. 573-581, 1991.

2. In a repeated game, of which

	L	R
T	$1,1$	$0,0$
B	$0,0$	$-1,4$

   is the stage game, the payoff profile $(3/4, 3/4)$.

   (a) Is feasible and enforceable.

   $\surd$ Is feasible but not enforceable.

   (c) Is enforceable but not feasible.

   (d) Is neither feasible nor enforceable.

> **Explanation.** The set of feasible payoff profiles is given by the convex hull of the four payoff profiles that correspond to the four action profiles. The payoff profile $(3/4, 3/4)$ can be realised with, e.g., row playing T 75% of the time, while col playing L all of the time. So the payoff profile $(3/4, 3/4)$ is feasible with strategy profile $s = ((3/4, 1/4), (1, 0))$.
>
> The area of enforceable payoffs is given by combinations that supersede the minmax values (punishment values) of both players. The minmax value of row is 0 (col plays R throughout). The minmax value of col is $4/5$ (row plays T 80% of the time). The payoff profile $(3/4, 3/4)$ cannot be enforced: row may just as well play any action since its minmax value supersedes $3/4$.

3. An advantage of cumulative payoff matching (CPM) over average payoff matching (APM) is the following.

   (a) CPM does not lock in as fast as APM does.

   (b) CPM tends to converge faster than APM.

   $\surd$ In the long run, the exploitation rate (e.r.) of CPM is as high as, and usually higher than, the e.r. of APM.

   $\surd$ In the long run, the e.r. of CPM approaches one, while the e.r. of APM usually does not.

> **Explanation.** (d): $P\{$ any sub-dominant action will be played $\} \rightsquigarrow 0$ (slides; Beggs, 2005). This makes CPM an interesting algorithm from a theoretical point of view
>
> (a) is unknown, let alone it would be an explanation for anything. (b) is false: CPM sometimes converges extremely slowly because it tends to stick to sub-optimal actions very long if those actions pay out relatively well. The slow convergence of CPM makes it uninteresting from a practical point of view. (c) is it is implied by (d) hence weaker. It therefore drops out as a correct answer (see rules on MC questions, part "Every multiple-choice question has exactly one correct answer").
>
> Nevertheless, (c) is accepted as correct as well because the distinction between (c) and (d) is considered too subtle to make a difference.

4. ϵ-Greedy reinforcement learning with finitely many actions.

   (a) Some actions are explored finitely many times.

   (b) Some actions are explored finitely many times a.s.

   (c) Every action is explored infinitely many times.

   $\surd$ Every action is explored infinitely many times a.s.

> **Explanation.** From the second Borel-Cantelli lemma it follows that, with probability one, every action is explored infinitely many times. So it is not certain every action is explored infinitely many times. So it is possible that one or more actions are explored only finitely many times. The probability of such an event, however, is zero.

5. The following game is played with regret matching with initial action profile (T, L). Give the action and the cumulative regrets of the column player at the 5th round.

	L	R
T	$1, 4$	$3, 1$
B	$2, 0$	$0, 5$

   (a) L, $(2, 1)$.

   $\surd$ R, $(2, 1)$.

   (c) L, $(1, 2)$.

   $\surd$ R, $(1, 2)$.

> **Explanation.** (d) See the slides on no-regret and/or Ch. 2 of "Strategic Learning an its Limits (H. Peyton Young, 2004). On second thoughts, (b) is accepted as well because it is not clear for everyone that the convention is to put row first in action profiles, payoff profiles, and, in this case, regret profiles, and my opinion is that this question may not take this convention as a prerequisite.

$n:$	1	2	3	4	5
$a_{\text{row}}:$	T	B	B	T	T
$a_{\text{col}}:$	L	L	R	R	$\mathbf{R}$
$x_{\text{row}}:$	1	2	0	3	3
$x_{\text{col}}:$	4	0	5	1	1
$\Delta r(T):$	0	-1	3	0	0
$r(T):$	0	-1	2	2	2
$\Delta r(B):$	1	0	0	-3	-3
$r(B):$	1	1	1	-2	-5
$\Delta r(L):$	0	0	-5	3	3
$r(L):$	0	0	-5	-2	$\mathbf{1}$
$\Delta r(R):$	-3	5	0	0	0
$r(R):$	-3	2	2	2	$\mathbf{2}$

6. What is the benefit of ϵ-Greedy regret matching, compared to ordinary regret matching?

   $\surd$ Opponent behaviour is irrelevant.

   (b) Regret matching is done off-policy.

   (c) Regret matching is done only $\epsilon\%$ of the time, not all of the time.

   (d) For all $\delta > 0$ there exists an $\epsilon > 0$ such that ϵ-Greedy regret matching yields at most δ regret.

Explanation. Knowledge of opponent's actions is not needed, because regrets are computed when experimenting. Cf. slides regret matching and Peyton Young's monograph, p. 23. The second and third answer are actually false: with ϵ-Greedy regret matching, regret matching is done most of the time, based on occasional experiments. The last answer merely quotes the definition of ϵ-Greedy regret matching.

7. The maximum number of ESSs in a symmetric normal form game with n actions and k Nash equilibria, all mixed, is

   $\surd$ 1
   (b) k
   (c) 2^n
   (d) $2^n - 1$

Explanation. Every ESS is a NE. If an ESS is mixed, it is unique.

8. Consider the empirical frequencies in fictitious play on a normal form game with mixed equilibria only.

   (a) These do not converge.
   (b) These converge to an equilibrium.
   (c) These converge, but not necessarily to an equilibrium.
   $\surd$ Another answer.

Explanation. For some but not all games with mixed equilibria only, the empirical frequencies converge. For example, for matching pennies they converge, but for the Shapley game they do not. It is a theorem that, if the empirical frequencies converge in FP, they converge to a NE. (The actually played responses per stage need not be NE of the stage game, however, as is the case with, for instance, matching pennies.)

9. Smoothed fictitious play with smoothing parameter $\gamma > 0$, exponentiated regret matching with exponent $a \geq 0$.

   i) Behave identically when $\gamma \downarrow 0$ and $a \to \infty$.
   ii) Behave identically when $a = 0$ and $\gamma \to \infty$.

   (a) None.
   (b) Only *i)*.
   (c) Only *ii)*.
   $\surd$ Both.

Explanation. When $\gamma \downarrow 0$ and $a \to \infty$, then both converge to fictitious play. When $a = 0$ and $\gamma \to \infty$, then exponentiated regret matching plays randomly, and smoothed fictitious play converges to random play.

10. *i)* G is a game with a NE that is not a NSS.
    ii) H is a game with an ESS that is a non-strict NE.

    Which of these statements is possible?

    $\surd$ Both.
    (b) Only *i)*.
    (c) Only *ii)*.
    (d) None.

Explanation. Strict Nash ⇒ ESS ⇒ NSS ⇒ Nash, and none of the implications can be reversed.

11. Determine the ESSs of

	L	R
L	$1,1$	$7,2$
R	$2,7$	$3,3$

.

(a) There are none.

(b) $\{(2/7, 5/7)\}$

√ $\{(4/5, 1/5)\}$

(d) Another answer.

Explanation. This game has two pure equilibria $(1,0)$, $(0,1)$ and one mixed equilibrium $(1/3, 2/3)$. The corresponding mixed strategies are $p = (4/5, 1/5)$ and $q = (4/5, 1/5)$. (See the game theory slides on how to determine mixed equilibria.) The mixed equilibrium is also symmetric. Only symmetric equilibria are candidates for ESSs, so (p, q) is the only equilibrium to consider.

Since $p = (4/5, 1/5)$ is fully mixed, every response q is a best response to p (again, see the game theory slides to see why the latter is true):

$$\text{for all } q : q^T Ap \geq p^T Ap. \text{ In particular for all } q : q^T Ap = p^T Ap.$$

So the first condition of an ESS is violated. We'll have to verify the second condition of an ESS:

$$\text{for all } q \neq p : q^T Aq < p^T Aq.$$

Let $q = (y, 1-y)$, $y \neq 4/5$, be arbitrary. Then

$$p^T Aq = \begin{pmatrix} 4/5 & 1/5 \end{pmatrix} \begin{pmatrix} 1 & 7 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} y \\ 1-y \end{pmatrix} = \frac{31}{5} - 5y,$$
$$q^T Aq = \begin{pmatrix} y & 1-y \end{pmatrix} \begin{pmatrix} 1 & 7 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} y \\ 1-y \end{pmatrix} = 3 + 3y - 5y^2.$$

It is easy to verify that $q \neq p \;\Rightarrow\; q^T Aq < p^T q$, because

$$\begin{aligned} p^T Aq - q^T Aq &= \left(4\frac{1}{4} - 4y\right) - \left(4 - 2y - 4y^2\right) \\ &= \frac{16}{5} - 8y + 5y^2 \end{aligned}$$

which is positive on $[0,1]\backslash\{\frac{4}{5}\}$. So the second condition of an ESS is met. It follows that (p, p) is a symmetric equilibrium that corresponds to an ESS p.

12. Given the normal form game with symmetric payoff matrix

$$M = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix},$$

infer the corresponding replicator equation for the row player in terms of x, where $x \in [0,1]$ represents the proportion of playing the first action.

(a) $\dot{x} = x(1 - 3x + 2x^2)$, equilibrium at $x = 1/3$.

√ $\dot{x} = x(1 - 3x + 2x^2)$, equilibrium at $x = 1/2$.

(c) $\dot{x} = x(1 + 3x - 2x^2)$, equilibrium at $x = 1/3$.

(d) $\dot{x} = x(1 + 3x - 2x^2)$, equilibrium at $x = 1/2$.

Explanation. Fitness of row given $p = (x, 1 - x)$:

$$f_{\text{row}} = (1\ 0)Ap = \begin{pmatrix}1 & 0\end{pmatrix} \begin{pmatrix}1 & 3\\ 2 & 2\end{pmatrix} \begin{pmatrix}x\\ 1-x\end{pmatrix} = \begin{pmatrix}1 & 0\end{pmatrix} \begin{pmatrix}3-2x\\ 2\end{pmatrix} = 3 - 2x.$$

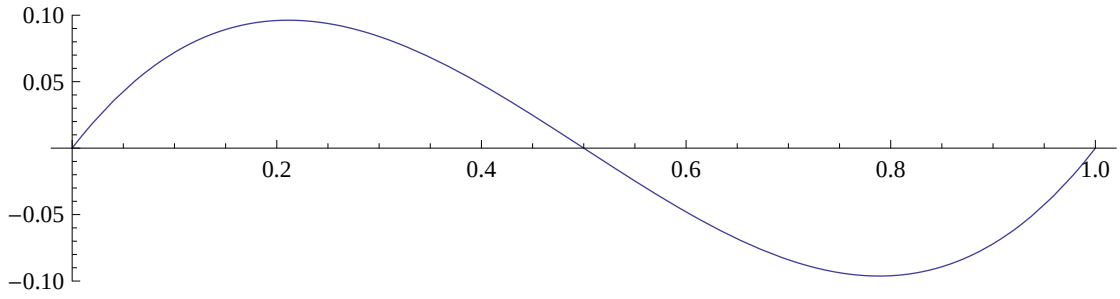
Average fitness given $p = (x, 1 - x)$:

$$\bar{f} = pAp = \begin{pmatrix}x & 1-x\end{pmatrix} \begin{pmatrix}1 & 3\\ 2 & 2\end{pmatrix} \begin{pmatrix}x\\ 1-x\end{pmatrix} = \begin{pmatrix}x & 1-x\end{pmatrix} \begin{pmatrix}3-2x\\ 2\end{pmatrix} = 2 + x - 2x^2.$$

The replicator equation stipulates that the change in proportion of a species over time is determined by the difference between its fitness and the average fitness:

$$\begin{aligned}\dot{x} &= x(f_{\text{row}} - \bar{f})\\ &= x(3 - 2x - (2 + x - 2x^2))\\ &= x(1 - 3x + 2x^2).\end{aligned}$$

Plot:



Unstable (but trivial) equilibria at $x = 0$ and $x = 1$, Stable (and non-trivial) equilibrium at $x = 1/2$.

Document last modified on Friday 24th May, 2019 at 17:51.