CIS603 – Multi-agent Systems

Mid-term Exam **Solutions**

30 October 2013

This exam is closed book, closed notes, closed electronic devices, and open brain. The exam consists of 12 really really really easy questions or groups of questions. The point value for each question or group of questions is provided. You will have 1.4 hours to take the exam, at which point you will need to turn it in. Thus, if a particular question is taking a particularly long time, skip it and return back to it later. Show your work where applicable, as partial credit is possible. Only legible answers and work can be given credit.

1. (24 points – 3 points each) Indicate whether each statement is *false* (i.e., it contains some untruth) or *true* (otherwise).

   a. (**<u>True</u>** or **False**) In traditional game theory, all players are assumed to always act rationally.

   b. (**<u>True</u>** or **False**) Every normal-form game has at least one Nash equilibrium.

   c. (**True** or **<u>False</u>**) An algorithm is collectively stable if it cannot be invaded by ~~tit-for-tat~~ any ~~other~~ algorithm.

   d. (**True** or **<u>False</u>**) The Nash bargaining solution is a Pareto optimal solution that maximizes the ~~sum of the players' payoffs~~ product $(P_1 - V_1^{mm})(P_2 - V_2^{mm})$.

   e. (**True** or **<u>False</u>**) In general-sum games, it is a Nash equilibrium when both players play their maximin strategies.

   f. (**<u>True</u>** or **False**) Every finite extensive-form game can be converted to a normal-form game.

   g. (**True** or **<u>False</u>**) A single agent playing tit-for-tat can invade a population of agents that always defect in the prisoners' dilemma when the discount factor is sufficiently high. *But a group of TFT agents can if they interact among each other enough.*

   h. (**True** or **<u>False</u>**) Tit-for-tat is the best strategy in the prisoners' dilemma. *There is no best strategy against an unknown associate.*

	d	**c**
d	1, 1	5, 0
c	0, 5	3, 3

2. (4 points) How will *rational* (self interested) agents play a repeated prisoners' dilemma when it is known that the game will last exactly 10 moves? Explain your answer.

   ANSWER:
   Always defect (AD). Backward reasoning: Best strategy in last move for both agents is to always defect. Thus, there is no incentive to cooperate in the second to last move. Same reasoning can be applied to third to last move, etc. The process backs up to the first move.

3. (3 points) What is an agent's best strategy (with respect to maximizing its own expected payoff) in a one-shot prisoners' dilemma played against an associate with an unknown strategy? Briefly explain your answer.

   ANSWER:
   Always defect (AD) as it is strategically dominant.

4. (5 points) What is an agent's best strategy (with respect to maximizing its average *per move* payoff) in a repeated prisoners' dilemma with a discount factor near 1 when playing against an associate with an unknown strategy? Prove your answer.

   ANSWER:
   There is no best strategy against an unknown opponent. Only if the opponent is known (or assumed) can we play best response. For example, if the opponent plays tit-for-tat (TFT) it's best to always cooperate (or play TFT). If the opponent is always defect (AD), it's best to AD in turn.

5. (5 points) In an repeated prisoners' dilemma with a discount factor near 1, suppose that the strategy of the first agent is to play tit-for-tat, while the strategy of the second agent is to always defect. Is this a Nash equilibrium? Why or why not.

   ANSWER:
   No, it is not a NE, because AD is not best response to TFT, and TFT is not best response to AD.

6. (8 points) What are the row and column player's maximin strategies and values in the following matrix game? Show your work.

	c	**d**
a	3, -1	2, 2
b	2, 4	4, 3

ANSWER:

Row player: let p be the probability of row player playing a.
$U(a|c) = 3p + 2(1-p) = 3p + 2 - 2p = p + 2$
$U(a|d) = 2p + 4(1-p) = 2p + 4 - 4p = 4 - 2p$
$U(a|c) = U(a|d)$
$p + 2 = 4 - 2p$
$3p = 2$
$p = 2/3$

Maximin strategy for row player: (a: 2/3, b: 1/3)
Maximim value for row player (or value of the game): p + 2 = 2/3 + 2 = 8/3

Column player: let q be the probability of column player playing c.
$U(c|a) = -q + 2(1-q) = -q + 2 - 2q = 2 - 3q$
$U(c|b) = 4q + 3(1-q) = 4q + 3 - 3q = 3 + q$
$U(c|a) = U(c|b)$
$2 - 3q = 3 + q$
$-1 = 4q$
$q<0$ *(not possible by definition of probability)*

Maximin strategy for column player: (c: 0, d: 1)
Maximim value for column player: 2

7. (6 points) What are the pure strategy Nash equilibria of the following two matrix games?

	c	**d**
a	2, 2	0, 3
b	3, 0	4, 4

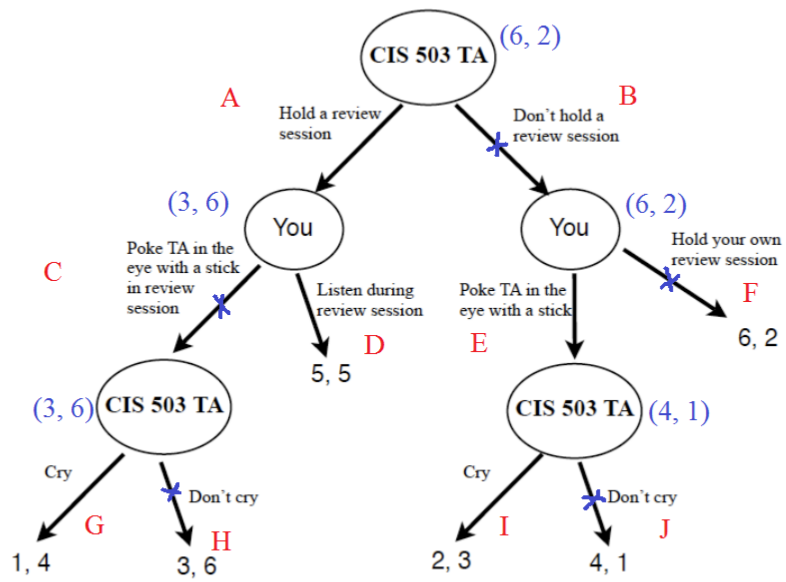
	c	**d**
a	2, 4	4, 3
b	3, 0	2, 2

ANSWER:
First game: (4, 4) is NE.
Second game: No pure strategy NE exists.

8. (5 points) What is/are the pareto optimal (PO) solution(s) of the following matrix game?

	a	**b**
c	3, 3	2, 2
d	5, 0	4, 2

ANSWER:
(3, 3), (5, 0) and (4, 2) are Pareto optimal.



9. (12 points) Convert the above extensive-form game to normal form. Payoffs to the TA are listed first. Determine the pure-strategy Nash equilibria of the game and indicate which of these Nash equilibria are sub-game perfect and which are not.

ANSWER:

Pure strategies for TA: AGI, AHI, AGJ, AHJ, BJI, BHI, BGJ, BHJ.
Pure strategies for You: CE, CF, DE, DF.

	CE	CF	DE	DF
AGI	1, 4	1, 4	5, 5	5, 5
AHI	3, 6	3, 6	5, 5	5, 5
AGJ	1, 4	1, 4	5, 5	5, 5
AHJ	3, 6	3, 6	5, 5	5, 5
BGI	2, 3	6, 2	2, 3	6, 2
BHI	2, 3	6, 2	2, 3	6, 2
BGJ	4, 1	6, 2	4, 1	6, 2
BHJ	4, 1	6, 2	4, 1	6, 2

(AGI, DE), (AGJ, DE), (BGJ, CF), (BHJ, CF), (BGJ, DF), (BHJ, DF) are NE.
(BHJ, CF) is sub-game perfect NE.

10. (8 points) Consider the following payoff matrix for a two-player normal form game:

	d	**e**
a	3, 6	3, 4
b	2, 2	5, 5
c	0, 3	6, 1

A. (4 points) Suppose the column player plays action **d** with probability $\frac{1}{2}$. How many (a) pure-strategy best-responses, and (b) how many mixed-strategy best-responses are there for the row player?

ANSWER:

$$U(a|(\frac{1}{2},\frac{1}{2})) = U((1,0,0)|(\frac{1}{2},\frac{1}{2})) = 3(\frac{1}{2}) + 3(\frac{1}{2}) = 3$$
$$U(b|(\frac{1}{2},\frac{1}{2})) = U((0,1,0)|(\frac{1}{2},\frac{1}{2})) = 2(\frac{1}{2}) + 5(\frac{1}{2}) = 3.5$$
$$U(c|(\frac{1}{2},\frac{1}{2})) = U((0,0,1)|(\frac{1}{2},\frac{1}{2})) = 0(\frac{1}{2}) + 6(\frac{1}{2}) = 3$$

There is one pure-strategy best response, namely b.

Since there is only one pure-strategy best-response to $(\frac{1}{2},\frac{1}{2})$, then there is no mixed-strategy best-response to $(\frac{1}{2},\frac{1}{2})$. In fact, every mixed-strategy (by the row player) will gain strictly less than 3.5. Take the strategy (p_a, p_b, p_c) which means playing a with probability p_a, b with probability p_b and c with probability p_c. We have $p_a + p_b + p_c = 1$. Then:

$$U((p_a,p_b,p_c)|(\frac{1}{2},\frac{1}{2})) = p_a.U(a|(\frac{1}{2},\frac{1}{2})) + p_b.U(b|(\frac{1}{2},\frac{1}{2})) + p_c.U(c|(\frac{1}{2},\frac{1}{2})) \quad (1)$$
$$= (3)p_a + (3.5)p_b + (3)p_c \quad (2)$$

Note that this amount will be strictly less than 3.5 unless $p_b = 1$ (i.e. $p_a = 0$ and $p_c = 0$). If you try different values for p_a,p_b, and p_c (given they add up to 1), you will find that the higher p_b is, the higher the overall utility is. Hence, $p_b = 1$ gives the highest utility.

B. (4 points) Repeat for the case in which the column player plays action **d** with probability $\frac{2}{3}$.

ANSWER:

$$U(a|(\frac{2}{3},\frac{1}{3})) = U((1,0,0)|(\frac{2}{3},\frac{1}{3})) = 3(\frac{2}{3}) + 3(\frac{1}{3}) = 3$$
$$U(b|(\frac{2}{3},\frac{1}{3})) = U((0,1,0)|(\frac{2}{3},\frac{1}{3})) = 2(\frac{2}{3}) + 5(\frac{1}{3}) = 3$$
$$U(c|(\frac{2}{3},\frac{1}{3})) = U((0,0,1)|(\frac{2}{3},\frac{1}{3})) = 0(\frac{2}{3}) + 6(\frac{1}{3}) = 2$$

There are two pure-strategy best-responses, namely a and b.

Since there are two pure-strategy best-responses to $(\frac{2}{3},\frac{1}{3})$, then there is an infinite number of mixed-strategy best-responses to $(\frac{2}{3},\frac{1}{3})$. In fact, every mixed-strategy in which $p_c = 0$ (prob. of playing c is 0) will gain exactly 3:

$$(3)p_a + (3)p_b + (2)p_c = (3)p_a + (3)p_b = 3(p_a + p_b) = 3$$

11. (10 points) For the following repeated matrix game with a high discount factor (near 1), describe a pair of trigger strategies that form a Nash equilibrium of the repeated game in which the row player gets an average payoff (per round) of 3.5 and the column player receives an average payoff (per round) of 6.

	c	**d**
a	5, 2	3, 1
b	4, 4	2, 10

ANSWER:
Each player proposes the following offer:

(a) If you cooperate with me so that we alternate between the profiles (a,c) and (b,d), I will promise to do my part truthfully.

(b) If you deviate, I will punish you and try to minimize your maximum payoff (i.e. play minimax). In other words, the row player punishes by playing always a (hence, giving a maximum of 2 for the column player), and the column player punishes by playing always d (hence, giving a maximum of 3 for the row player).

Note: there can be different forms of punishments. However, the punishment needs to give the other player less than what she gets when she cooperates.

Briefly state why this pair of trigger strategies constitutes a Nash equilibrium.

ANSWER:

- When the column player plays c in the odd rounds and d in the even rounds, the row player has two options: (1) to gain an average of 3.5 by cooperating (i.e. playing a in the odd rounds and b in the even rounds), or (2) to gain a maximum average of 3 by deviating and getting punished by the column player.
- When the row player plays a in the odd rounds and b in the even rounds, the column player has two options: (1) to gain an average of 6 by cooperating (i.e. playing c in the odd rounds and d in the even rounds), or (2) to gain a maximum average of 2 by deviating and getting punished by the column player.

Hence, cooperating is a best response for both agents.

Is the average payoff obtained by the agents Pareto optimal?

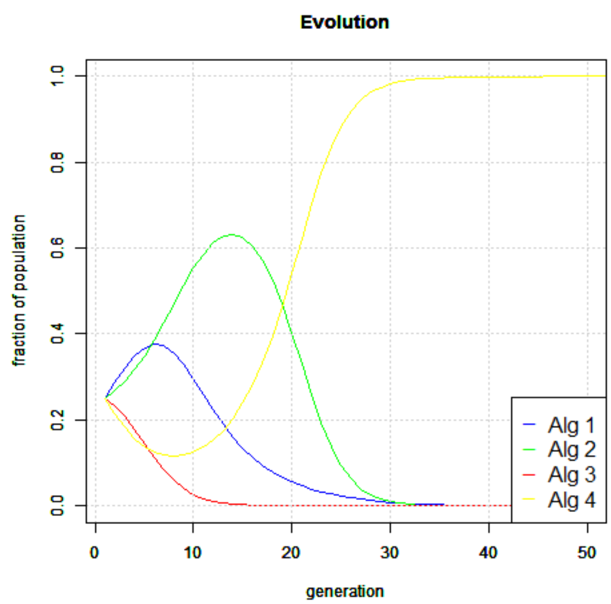
ANSWER:
Yes. The payoff profile $(3.5,6)$ is on the line connecting $(5,2)$ and $(2,10)$, which happens to be on the dominant side of the convex hull. Hence, it is Pareto optimal.

12. (10 points) The table below shows the results of a round-robin tournament played by four different algorithms. Each cell of the table shows the average payoff of the algorithm listed at the beginning of the row against the algorithm listed at the top of the column in a repeated game. Clearly, **Alg 1** is the winner of the round robin tournament. Which algorithm do you think would win an evolutionary tournament run using the replicator dynamic (you don't need to actually simulate this tournament)? Give reasons for why you think that algorithm would win?

	Alg 1	**Alg 2**	**Alg 3**	**Alg 4**	**Average**
Alg 1	3	3	6	4	**4.00**
Alg 2	4	4	4	3	**3.75**
Alg 3	2	2	5	4	**3.25**
Alg 4	2	5	0	5	**3.00**

ANSWER:

The winner is Alg 4. If you are skeptic, check the following figure.



However, one can still anticipate the probable winner without even running the evolutionary tournament. Note that for an algorithm X to win in the evolutionary tournament, it likely needs to be collectively stable (with respect to the participating algorithms) i.e. a group of agents playing algorithm X cannot be invaded by a single agent playing a different algorithm.

Why are "winning the evolutionary tournament" and "being collectively stable" related? Suppose an algorithm X is the winner of the round-robin tournament. Hence, at first, X is the leading algorithm in the evolutionary tournament. If X is not collectively stable, then there exists another algorithm Y that will invade X, and then Y becomes the leading algorithm. If Y itself is not collectively stable, then there is another algorithm Z that will invade Y, and then Z becomes the leading algorithm. One the other hand, if algorithm Y is collectively stable, then no algorithm can invade it and Y keeps the lead and win.

How do we find if an algorithm X is collectively stable or not? Well, we learned from Axelrod's book, that X is collectively stable iff there is no other algorithm Y that can invade X. How do we know if an algorithm Y can invade algorithm X? We compare the payoffs gained by Y when playing against X with the payoffs gained by X when playing against itself. If the former is higher, then Y can invade X and hence X is not collectively stable. If there is no algorithm that can invade X then X is likely to win the evolutionary tournament.

In this question, Alg 1 has the lead at first. However, Alg 1 is not collectively stable since Alg 2 can invade it (Alg 2 gains 4 against Alg 1, while Alg 1 gains 3 against itself). Then, Alg 2 has the lead. However, Alg 2 itself is not collectively stable, since Alg 4 can invade it (Alg 4 gains 5 against Alg 2, while Alg 2 gains 4 against itself). Then, Alg 4 has the lead. Alg 4 is collectively stable (Alg 4 gains 5 against itself, no other algorithm gets more than that when playing against Alg 4). Hence, Alg 4 is the winner.