Equilibrium in Multi-stage Game Playing with Dynamic Programming

Beidi Qiang

Southern Illinois University Edwardsville, Edwardsville, IL, USA

Abstract In this paper, the problem of multi-stage game playing is studied in a decision theory framework. The conditional strategies of payers are examined under the assumption of perfect information about previously played stages. The equilibrium is defined in analogy to the Nash equilibrium and the existence of such equilibrium is discussed mathematically through the idea of dynamic programming and backward induction. An example in a two-players zero-sum game setting is analyzed with details and the solution of the equilibrium strategy is presented.

Keywords Multi-stage game, Dynamic game, Nash equilibrium, Dynamic programming

1. Introduction

In game theory, Nash equilibrium is one of the most well-studied solutions to a single-stage non-cooperative game. An equilibrium in the game is defined as no player can do better by unilaterally changing his strategy. It is proved that if mixed strategies are allowed, then a game with a finite number of players and finitely many pure strategies has at least one Nash equilibrium [1]. The problem of finding equilibrium in various single-stage game settings has been studied in the literature.

In reality, dynamic game playing over time is more complex than one game. Instead, decision-makers are expected to play one game followed by another, or maybe several other games. It is tempting to view each game independently and assume players do not make any attempt to link them. If each player treats every game independently, equilibrium can be studied for each of the stage games in turn. However, this approach ignores any real issues concerning sequential behavior, if we consider the games played sequentially as stages of a larger game and players can observe the outcomes of each game before another game is played. If players are rational and forward-looking, they should view the sequence of decision-making as part of one large game. And if they do, we expect their actions to be dependent on the outcome of the previous games.

The existence and solutions of equilibrium for dynamic games are well-studied in the literature. The existence of sub-game equilibria in simultaneous move games with finite actions was discussed by Fudenberg and Levine 1983

[2]. Kreps and Wilson [3] proposed the concept of sequential equilibrium and Myerson [4], in his key paper in 1986, discussed the sequential communication equilibrium in multi-stage games with communication mechanisms that can be implemented by a central mediator. There has been a surge of discussions in multistage game playing with various designs of dynamic mechanisms and information systems. For example, Renault, Solan, and Vieille [5] studied a dynamic model of information provision with a greedy disclosure policy, and He and Sun [6] discussed the existence of subgame-perfect equilibria in dynamic games with simultaneous moves.

In this paper, we discuss details of conditional decision-making in a multi-stage game with players choosing their actions simultaneously at each stage. Games at each stage are played sequentially, with actions and information from previous games unfolding, but the payoffs were delayed until the sequence ends. These games form a general class of dynamic games with many applications in economics and political science. We define the settings of a multi-stage game and its equilibrium in section 2. The existence of equilibrium in this game setting is discussed through the idea of backward induction in section 3. An example of a simple game with two players and two stages is presented in section 4. Concluding remarks are discussed in section 5.

2. The Setting of a Multi-Stage Game

Consider a multi-stage game with n players and finite K stages. At each stage, player i, has a finite strategy space $\mathcal{A}_i^k$. Let $a_i^k \in \mathcal{A}_i^k$ denote the pure strategy chosen by player i at stage k. Then $a^k = (a_1^k, a_2^k, \cdots, a_n^k)$ is

the profile of strategies used by each player at stage k, where $a^k = (a_1^k, a_2^k, \cdots, a_n^k) \in \mathcal{A}^k = \mathcal{A}_1^k \times \mathcal{A}_2^k \times \cdots \times \mathcal{A}_n^k$. Assume the game at each stage is a simultaneous move game, i.e a normal-form game.

At stage k, the outcome for player i is denoted by $x_i^k \in \mathcal{X}_i^k$ and the combined outcome for all players is denoted by vector $x^k = (x_1^k, \cdots, x_n^k) \in \mathcal{X}^k = \mathcal{X}_a^k \times \cdots \times \mathcal{X}_n^k$. Assume the combined outcome at stage k comes from a joint distribution that is known given the joint choice of actions, i.e. $x^k \sim F_{a^k}^k$, where $F_{a^k}^k$ depends on the joint action a^k.

Assume each game is played in a distinct period and also that after each stage is completed, all players observe the outcome of that stage and all outcomes gained before that stage, i.e. $x^1, \cdots, x^{k-1}$ are observed for each player at stage k. It is plausible that the action of each player at stage k will depend on the observed outcome at previous stages. In this setting, the outcomes of previous stages become common knowledge, and players are allowed to make conditional strategies for future games based on past outcomes.

2.1. The Payoff

The games are played sequentially, and the total payoffs from the sequence of games will be evaluated using the sequence of outcomes at the end of the games, i.e. after stage K. Suppose each player receives a real number payoff or reward. The reward function g_i for player i is defined as a function of the outcomes of all K stages, i.e. $g_i : (\mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^k) \to \mathbb{R}$. The joint reward function for all players is then defined as $G : (\mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^k) \to \mathbb{R}^n$, where $G = (g_1, \cdots, g_n)$. With the reward function defined, the expected reward of each player can be calculated when the joint strategies at all stages are given.

Given the joint actions at all stages $a^1, \cdots, a^K$, the expected reward for player i is defined as

$$\tilde{r}_i(a^1, \cdots, a^K)$$
$$= E[g_i(x^1, \cdots, x^k) \mid a^1, \cdots, a^K]$$
$$= \int_{\mathcal{X}^1} \cdots \int_{\mathcal{X}^{K-1}} \int_{\mathcal{X}^K} g_i(x^1, \cdots, x^k)$$
$$F_{a^K}^K(dx^K) F_{a^{K-1}}^{K-1}(dx^{K-1}) \cdots F_{a^1}^1(dx^1).$$

2.2. Mixed Strategy

In a general setting, at a given stage, each player is allowed to have mixed strategy, i.e. players choose actions based on a probability distribution of possible pure strategies at the stage. Consider the decision function for player i at stage k, δ_i^k, as a function from the product space of the previous outcome space $\mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^{k-1}$ to the probability space of strategies at stage k, $\mathcal{P}(\mathcal{A}_i^k)$. And the joint decision function for all players at stage k is then denoted by $\delta^k = (\delta_1^k, \cdots, \delta_n^k)$, where $\delta^k : \mathcal{X}^1 \times \cdots \times \mathcal{X}^{k-1} \to \mathcal{P}(\mathcal{A}^k)$.

Assume the joint reward $G = (g_1, \cdots, g_n)$ at the end of the game is defined the same as above. Given the joint mixed strategies at all stages, $\delta = (\delta^1, \cdots, \delta^K)$, the expected reward for player i is can be calculated as

$$r_i(\delta) = E[g_i(x^1, \cdots, x^k) \mid \delta^1, \cdots, \delta^K]$$
$$= \int_{\mathcal{A}} \int_{\mathcal{X}^1} \cdots \int_{\mathcal{A}^K} \int_{\mathcal{X}^K} g_i(x^1, \cdots, x^k)$$
$$F_{a^K}^K(dx^K) \delta^K(da^K, x^1, \cdots, x^{K-1}) \cdots F_{a^1}^1(dx^1) \delta^1(da^1).$$

2.3. Definition of Equilibrium

In a single-stage game, the Nash equilibrium is achieved when no player can increase his own expected payoff by switching his strategy. A collection of mixed strategies for all players, $\boldsymbol{\delta}^*$, is considered to be the equilibrium strategies if for all $i = 1, \ldots, n$,

$$r_i(\boldsymbol{\delta}^*) = \max_{\delta_i} r_i(\boldsymbol{\delta}_{-i}^*, \delta_i)$$

where $(\delta_{-i}^*, \delta_i)$ denotes a new set of mixed strategies when only player i switched from mixed strategy δ_i^* to δ_i.

Due to the multi-stage nature of the game, the decisions for all players are made stage by stage based on the previously observed outcomes. The equilibrium joint strategies of the game is thus defined sequentially as no players can be better off by switching to another strategy at each stage conditionally on observed previous outcomes. The joint mixed strategies for all players at all stages $\delta^* = (\delta^{1*}, \cdots, \delta^{K*})$, is considered to be an equilibrium if at any stage k given outcomes from stage 1 to $k-1$, for all player $i = 1, \cdots, n$. The expected reward of player i at the equilibrium satisfies the following,

$$r_i(\delta^{k*}, \delta^{(k+1)*}, \cdots, \delta^{K*} \mid x^1, \cdots, x^{k-1})$$
$$= \max_{\delta_i^k} \{r_i[(\delta_{-i}^{k*}, \delta_i^k), \delta^{(k+1)*}, \cdots, \delta^{K*} \mid x^1, \cdots, x^{k-1}]\},$$

where $(\delta_{-i}^{k*}, \delta_i^k)$ denotes a new joint randomized decision at stage k when only player i switched from δ_i^{k*} to δ_i^k.

3. The Existence of Equilibrium

It has been shown that there is a Nash equilibrium for every single-stage finite game when mixed strategies are allowed. The proof of existence is outlined in Nash's original paper using Kakutani fixed-point theorem [7]. We now extend the discussion to a multi-stage game using the idea of dynamic programming and backward induction, which was developed by Richard E. Bellman [8]. The essence of dynamic programming is to simplify a complicated optimization problem by breaking it down into simpler sub-problems recursively. The nature of a multi-stage game allows us to break the game down into a sequence of sub-games over discrete time, with each sub-game as a finite non-cooperative game. Starting from the end of the game, we first examine the last stage and identify what strategy would be most optimal at that moment. We then backtrack and determine what to do at the second-to-last time the decision. This process continues backward until one has determined a sequence of optimal actions for every possibility at every point.

Consider the last stage, stage K, of the game when the outcomes of the previous stages are observed. The expected reward based on the joint decision of all players at the last stage, δ^K, is given by

$$r_i(\delta^K \mid x^1, \cdots, x^{K-1}) = \int_{A^K} \int_{X^K} g_i(x^1, \cdots, x^K)$$
$$F_{a^K}^K(dx^K)\delta^K(da^K, x^1, \cdots, x^{K-1}).$$

When the outcomes from stage 1 through $K-1$, i.e. $x^1, .. x^{K-1}$ are observed, the game at the last stage becomes a single stage game, which has an equilibrium as proved by Nash [7]. Denote the equilibrium decision at stage K based on previous observations as $\delta^{K*}(x^1, .. x^{K-1})$ and the equilibrium expected reward given outcomes from stage 1 to $K-1$ is given by $r_i(\delta^{K*} \mid x^1, .. x^{K-1})$. If the players follow the equilibrium strategy $\delta^{K*}(x^1, \cdots, x^{K-1})$ at stage K, then the expected reward can be written as

$$r_i(\delta^1, \cdots, \delta^{K-1}, \delta^{K*})$$
$$= E[g_i(x^1, \cdots, x^k) \mid \delta^1, \cdots, \delta^{K*}]$$
$$= \int_A \int_{X^1} \cdots \int_{A^K} \int_{X^K} g_i(x^1, \cdots, x^k)$$
$$F_{a^K}^K(dx^K)\delta^{K*}(da^K, x^1, \cdots, x^{K-1}) \cdots F_{a^1}^1(dx^1)\delta^1(da^1)$$
$$= \int_A \int_{X^1} \cdots \int_{A^{K-1}} \int_{X^{K-1}} r_i(\delta^{K*} \mid x^1, \cdots, x^{K-1})$$
$$F_{a^{K-1}}^{K-1}(dx^{K-1})\delta^{K-1}(da^{K-1}, x^1, \cdots, x^{K-2}) \cdots F_{a^1}^1(dx^1)\delta^1(da^1).$$

Assume all players follow joint strategy δ^{K*} at stage K, now consider stage $K-1$ given the information of previous outcomes up to stage $K-2$. The expected reward is then given as

$$r_i(\delta^{K-1} \mid \delta^{K*}, x^1, \cdots, x^{K-2})$$
$$= \int_{A^{K-1}} \int_{X^1} \cdots \int_{X^{K-1}} r_i(\delta^{K*} \mid x^1, \cdots, x^{K-1})$$
$$F_{a^{K-1}}^{K-1}(dx^{K-1})\delta^{K-1}(da^{K-1}, x^1, \cdots, x^{K-2}).$$

This again becomes a single-stage game, and Nash equilibrium exists. Denote the equilibrium decision at stage $K-1$ as $\delta^{(K-1)*}(x^1, \cdots, x^{K-2})$ and the equilibrium expected reward given outcomes from stage 1 to $K-2$ is given by $r_i(\delta^{(K-1)*}, \delta^{K*} \mid x^1, \cdots, x^{K-2})$. Repeat the process sequentially for stage $K-2, K-3, \cdots, 1$. We can then find an equilibrium decision for all stages, $\delta^* = (\delta^{1*}, \delta^{2*}, \cdots, \delta^{K*})$.

4. An Example

4.1. A Game with 2-Stages

Consider a simple game with $n = 2$ players and $K = 2$ stages. Players are non-cooperative and play simultaneously. Player A has 2 pure strategies to choose from at each stage, denoted by A_0 and A_1. Similarly, player B also has 2 possible pure strategies, B_0 and B_1. Assume the outcomes for player A and B are the same at each stage, and follow a distribution F_{ij}, if player A choose pure strategy i and player B choose pure strategy j at the stage, i.e. $x_a^k = x_b^k = x^k \sim F_{ij}$, $k = 1, 2$, $i, j = 0, 1$. Assume mixed strategies are allowed. Let α denote the probability of player A choosing pure strategy A_1 and β be the probability of player B choosing strategy B_1.

4.2. The Reward Function

Consider the zero-sum game, in which the sum of rewards for players A and player B equals zero. Consider the situation that the goal of player A in this game is to reach a threshold τ and the opponent B wants the opposite. Furthermore, assume the game stops whenever the threshold τ is achieved. Thus, the reward function at the end of the game is given by

$$g_a = 1_{\{x^1 + x^2 \geq \tau, x^1 < \tau\}} + 1_{\{x^1 \geq \tau\}},$$

where $g_b = -g_a$. The expected reward function is then given by

$$r_a(\alpha^1, \beta^1, \alpha^2, \beta^2) = P(x^1 + x^2 \geq \tau \mid \alpha_1, \beta_1, \alpha_2, \beta_2),$$

where $r_b = -r_a$.

4.3. Computing the Mixed Equilibria

We compute the mixed equilibrium strategy through backtracking. Start with stage 2 given the outcome in stage 1 and assume the threshold has not been achieved, i.e. x^1 is known and less than τ. The expected reward for player A is then given by

$$r_a(\alpha^2, \beta^2 \mid x^1) = P(x^2(\alpha^2, \beta^2) \geq \tau - x^1)$$
$$= \alpha^2 \beta^2 \bar{F}_{11}(\tau - x^1) + (1-\alpha^2)\beta^2 \bar{F}_{01}(\tau - x^1)$$
$$+ \alpha^2(1-\beta^2)\bar{F}_{10}(\tau - x^1) + (1-\alpha^2)(1-\beta^2)\bar{F}_{00}(\tau - x^1)$$

Now the problem at stage 2 becomes a single-stage zero-sum game, where the equilibrium can be obtained using the min-max decision rule (see, for example, [9]). The optimal strategies and the corresponding equilibrium at stage 2 are given by

$$\alpha^{2*} = \frac{\bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1)}{\bar{F}_{11}(\tau-x^1) + \bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1) - \bar{F}_{01}(\tau-x^1)},$$

$$\beta^{2*} = \frac{\bar{F}_{00}(\tau-x^1) - \bar{F}_{01}(\tau-x^1)}{\bar{F}_{11}(\tau-x^1) + \bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1) - \bar{F}_{01}(\tau-x^1)},$$

$$r_a^*(x^1) = \frac{\bar{F}_{11}(\tau-x^1)\bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1)\bar{F}_{01}(\tau-x^1)}{\bar{F}_{11}(\tau-x^1) + \bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1) - \bar{F}_{01}(\tau-x^1)}.$$

Now consider the first stage assuming the players follow the optimal strategies given above at the second stage.

$$r_a(\alpha^1, \beta^1, \alpha^2, \beta^2)$$
$$= P(x^1(\alpha^1,\beta^1) \geq \tau) + P(x^1 + x^2 \geq \tau, x^1 < \tau)$$
$$= \alpha^1\beta^1 H_{11} + (1-\alpha^1)\beta^1 H_{01} + \alpha^1(1-\beta^1)H_{10}$$
$$+ (1-\alpha^1)(1-\beta^1)H_{00},$$

where

$$H_{ij} = \int_{-\infty}^{\tau} r_a*(x^1)F_{ij}(dx^1) + \bar{F}_{ij}(\tau)$$
$$= \int_{-\infty}^{\tau} \frac{\bar{F}_{11}(\tau-x^1)\bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1)\bar{F}_{01}(\tau-x^1)}{\bar{F}_{11}(\tau-x^1) + \bar{F}_{00}(\tau-x^1) - \bar{F}_{10}(\tau-x^1) - \bar{F}_{01}(\tau-x^1)}$$
$$F_{ij}(dx^1) + \bar{F}_{ij}(\tau).$$

Now the game again becomes a single-stage zero-sum game, where the equilibrium can be obtained as follows

$$\alpha^{1*} = \frac{H_{00} - H_{10}}{H_{00} + H_{11} - H_{10} - H_{01}}, \quad (1)$$

$$\beta^{1*} = \frac{H_{00} - H_{01}}{H_{00} + H_{11} - H_{10} - H_{01}}, \quad (2)$$

$$r_a* = \frac{H_{00}H_{11} - H_{10}H_{01}}{H_{00} + H_{11} - H_{10} - H_{01}}. \quad (3)$$

4.4. Numerical Calculation and Simulation

Assume the outcomes for players A and B are the same at each stage and follow normal distributions (with different means and standard deviation 1). The distributions of the outcomes for each combination of actions are listed in table 1. Assume mixed strategies are allowed. The goal of player A in this game is to reach a threshold $\tau = 10$ and the opponent B wants the opposite. The game stops whenever the threshold $\tau = 10$ is achieved. The expected reward for player A is then given by $r_a(\alpha^1, \beta^1, \alpha^2, \beta^2) = P(x^1 + x^2 \geq 10 \mid \alpha_1, \beta_1, \alpha_2, \beta_2)$.

Table 1. The distributions for the random outcomes in each joint action

Actions	B_0	B_1
A_0	$N(\mu = 6, \sigma = 1)$	$N(\mu = 5, \sigma = 1)$
A_1	$N(\mu = 3, \sigma = 1)$	$N(\mu = 8, \sigma = 1)$

Following the calculation in the previous section, the expected reward for player A under equilibrium is 0.688 (equation (3)), and the equilibrium mixed strategies in stage 1 are given by $\alpha^{1*} = 0.199$ and $\beta^{1*} = 0.628$ (equation (1) and (2)). The integrations in the calculation of H_{ij}'s are numerical approximations since the integrals can not be obtained in closed form. The strategies of both players A and B in stage 2 will vary depending on the outcome from stage 1. Table 2 listed some examples of strategies played in the second stage. In the last example, since the goal has been achieved in stage 1, the game stopped without going into the second stage.

Table 2. Mixed strategies played in stage 2 under different stage 1 outcomes

Stage 1 outcome	Stage 2 Strategies
0	$\alpha_2 = 0.001, \beta_2 = 0.001$
1	$\alpha_2 = 0.008, \beta_2 = 0.008$
3	$\alpha_2 = 0.139, \beta_2 = 0.162$
5	$\alpha_2 = 0.259, \beta_2 = 0.621$
7	$\alpha_2 = 0.041, \beta_2 = 0.956$
9	$\alpha_2 = 0.001, \beta_2 = 0.999$
> 10	NA

A simulation study was performed, with 5000 such games played. In this simulation, player A reached the goal ($g_a = 1$) for 3438 games which are 68.94% out of 5000 games. The number of times and the proportion of each joint action played at each stage are listed in Table 3.

Table 3. The number of times each joint action played at each stage in 5000 simulated games

Joint Actions	Stage 1	Stage 2
A_1, B_1	622 (12.44%)	524 (10.48%)
A_1, B_0	381 (7.62%)	386 (7.72%)
A_0, B_1	2486 (49.72%)	2850 (57.00%)
A_0, B_0	1511 (30.22%)	1242 (24.84%)
No stage 2 played		16 (0.32%)

5. Discussion and Remarks

In this paper, we focus on conditional decision-making in a multi-stage game with players choosing their actions simultaneously in the absence of coalitions. We defined and showed the existence of the multi-stage equilibrium. We also provide algorithm steps toward the calculation of such equilibrium using dynamic programming. Opposed to the traditional non-cooperative game setting, the joint actions of a cooperative multi-stage game can also be studied. In fact, John von Neumann and Morgenstern [10] presented a theory that analyzes various of types of coalitions that can be formed. A prefigured discussion of backward induction in a 2-person zero-sum setting is provided in their book.

The multi-stage game setting discussed in this paper assumes a simultaneous game at each stage, i.e. player chooses their action without knowledge of the actions chosen by other players. Sequential behavior within a stage can also be added to the game setting, which allows the possibility that players to move in turns in each sub-game. The sequential game at each stage can be analyzed mathematically using combinatorial game theory. Equilibrium can be found using backward induction if perfect information is assumed [11]. Sequential playing simply adds an additional layer to the current dynamics.

Perfect information about outcomes at previous stages is assumed in the setting of a multi-stage game. However, it is possible to imagine a game where some stages are played before the outcomes of previous stages are observed. If no information from previous stages is observed at all, the game at each stage can be treated as no different from a single-stage simultaneous game. If some, but not all the information is revealed, the analysis can become complicated depending on the structure of the information. If not all players are conscious of the other players' actions and payoffs at each stage, the backward induction analysis will not be easily applied [12].

REFERENCES

[1] John F. Nash. Non-cooperative games. *Annals of Mathematics*, 54(2): 286–295, 1951.

[2] Drew Fudenberg and David Levine. Subgame-perfect equilibria of finite- and infinite-horizon games. *Journal of Economic Theory*, 31(2): 251–268, 1983.

[3] David M. Kreps and Robert Wilson. Sequential equilibria. *Econometrica,* 50(4): 863–894, 1982.

[4] Roger B. Myerson. Multistage games with communication. *Econometrica,* 54(2): 323–358, 1986.

[5] Jérôme Renault, Eilon Solan, and Nicolas Vieille. Optimal dynamic information provision. *Games and Economic Behavior,* 104:329–349, 2017.

[6] Wei He and Yeneng Sun. Dynamic games with (almost) perfect information. *Theoretical Economics,* 15(2): 811–859, 2020.

[7] John F. Nash. Equilibrium points in ¡i¿n¡/i¿-person games. *Proceedings of the National Academy of Sciences,* 36(1): 48–49, 1950.

[8] R. Bellman. *Dynamic Programming.* CreateSpace Independent Publishing Platform, 2017.

[9] M.J. Osborne and A. Rubinstein. *A Course in Game Theory.* The MIT Press. MIT Press, 1994.

[10] J. von Neumann, O. Morgenstern, H.W. Kuhn, and A. Rubinstein. *Theory of Games and Economic Behavior: 60th Anniversary Commemorative Edition.* Princeton Classic Editions. Princeton University Press, 2007.

[11] Charalambos D Aliprantis. On the backward induction method. *Economics Letters,* 64(2): 125–131, 1999.

[12] J. Watson. *Strategy: An Introduction to Game Theory.* W.W. Norton, 2008.