The Complexity of Verifying Finite State Equilibrium in Repeated Games with Imperfect Private Monitoring

by

Yu Qiu

(Under the Direction of Prashant Doshi)

Abstract

In game theory, we have three major categories of repeated games: repeated games with perfect monitoring, repeated games with imperfect public monitoring, and repeated games with imperfect private monitoring. Among them, repeated games with imperfect private monitoring are the hardest instances because players lack reliable information shared by all. If each player's action on the equilibrium path is given by an automaton with finite number of states, then the equilibrium in a repeated game with imperfect private monitoring is called a finite state equilibrium. Given k players, we apply the partially observable Markov decision process (POMDP) to a repeated game with imperfect private monitoring and analyze the complexity of verifying a finite state equilibrium. This framework is new and has significant applications in markets such as secret price cutting among firms.

Index words: Algorithmic Game Theory, Repeated Games with Imperfect Private Monitoring, Prisoners' Dilemma, Finite State Equilibrium, Complexity, POMDP

The Complexity of Verifying Finite State Equilibrium in Repeated Games with Imperfect Private Monitoring

by

Yu Qiu

B.Eng., Shanghai University of Electric Power, 2012

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2015

The Complexity of Verifying Finite State Equilibrium in Repeated Games with Imperfect Private Monitoring

by

Yu Qiu

Major Professor: Prashant Doshi

Committee: Liming Cai
Suchendra M. Bhandarkar

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

Dedication

This thesis is dedicated to my parents for their endless love, support and encouragement.

Acknowledgments

I thank my advisor Dr. Prashant Doshi for guiding my research direction. I greatly appreciate the opporunity to go to Japan and to talk with our research collaborators. I am very grateful to my committee members Dr. Liming Cai and Dr. Suchendra M. Bhandarkar for valuable suggestions and comments regarding my thesis.

I also express my gratitude toward Dr. Volker Kaibel, Dr. David Bremner and Dr. Endre Boros for their constructive discussions on some parts of my thesis.

Finally, I would like to thank all the people in the THINC lab, and the other faculty, staff, and friends who have given me help during my graduate studies.

Contents

Acknowledgements **v**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Contributions 2

1.2 Structure of Thesis 3

2 Background **4**

2.1 One-Shot Games and Repeated Games 4

2.2 Nash Equilibrium 6

2.3 Grim Trigger Strategy 7

2.4 Partially Observable Markov Decision Process 7

3 Related Work **10**

3.1 Repeated Games with Perfect Monitoring 10

3.2 Repeated Games with Imperfect Public Monitoring 13

3.3 Complexity Class of Finding Nash Equilibria in Different Types of Games . . 18

4 Finite State Equilibrium in Repeated Games with Imperfect Private Monitoring **20**

4.1 Repeated Games with Imperfect Private Monitoring 21

4.2 Partially Observable Markov Decision Process in Repeated Games with Imperfect Private Monitoring 22

4.3 An Example 24

5 Complexity Results **32**

5.1 Problem Instance 32

5.2 The Algorithm 33

5.3 Analysis of Complexity Class 46

6 Discussion and Future Work **49**

Bibliography **51**

List of Figures

3.1 Player i is using the grim trigger strategy for the prisoners' dilemma game with perfect monitoring 12

3.2 The grim trigger strategy for the prisoners' dilemma game with imperfect public monitoring 15

4.1 This path preautomaton corresponds to the grim trigger strategy. P and R are two automaton states denoting punishment and reward, respectively. . . 23

4.2 The initial value functions of player i 27

4.3 The value functions of grim trigger strategy and all other one-shot extensions of player i's path preautomaton 30

5.1 An example of a belief region of two states 39

5.2 An example of a belief region of three states 40

5.3 The sum of angles is 2π if the point P is in the plane of the 3D polygon and on the interior 41

List of Tables

2.1 The prisoners' dilemma game 5

3.1 The payoff table for the prisoners' dilemma game with the public monitoring 15

4.1 Joint distribution of private signals based on action profile (C,C) 25
4.2 Joint distribution of private signals based on action profile (D,C) 25
4.3 Joint distribution of private signals based on action profile (C,D) 25
4.4 Joint distribution of private signals based on action profile (D,D) 26

5.1 The basic solutions table for the first system of linear equations 36

6.1 A comparison in direct and iterative methods (algorithms) for solving linear equation systems 51

Chapter 1

Introduction

Game theory is the study of strategic decision making. More formally, it is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers. Game theory has been widely used in many areas: economics, political science and psychology. Also, it has come to play an increasingly important role in logic and computer science. Several logical theories have a basis in game semantics. In addition, computer scientists have used games to model interactive computations especially in the field of multi-agent systems [36].

In game theory, we can divide games into two categories: one-shot games and repeated games. In one-shot games, each player makes a decision only once and it carries no further repercussions (rewards or punishments), so each player will try her best to gain the greatest profit immediately, but in repeated games, strategies can be very different. All players learn how to play the game by playing it many times and updating their strategies based on current states and observations.

In repeated games with perfect monitoring, all players fully observe the action profile chosen at the end of each time step [34]. However, in repeated games with imperfect public monitoring, players no longer have the correct information about the past play. At the end of each time step, all players receive the same public signal which may be a noisy observation about other players' actions [34].

Repeated game with imperfect private monitoring is one of the newest and less studied instances in game theory. This model has broader and more significant applications in markets such as secret price cutting among firms [20]. In repeated games with imperfect private monitoring, we do not have information shared by all players, and each player only gets a noisy private signal that indicates other players' actions in each round [34]. This complicates computing the best response and realizing when the other players should be punished to sustain equilibrium.

If each player's action on the equilibrium path which is the game progression when players are using strategies in equilibrium, is given by an automaton with a finite number of states, then the equilibrium in a repeated game with imperfect private monitoring is called a finite state equilibrium [26].

Verifying a finite state equilibrium is not easy because the best response of player i is defined on an uncountable number of his beliefs of other players' states. Hence, computing value iteration for all beliefs in principle involves uncountably many calculations (i.e. one for each belief). However, the theory of partially observable Markov decision processes (POMDPs) [22] shows that we can confine attention to piecewise linear value functions, and for those particular value functions, the computation is performed using a finite number of steps as noted by Kandori and Obara [26].

1.1 Contributions

There are three main contributions of this thesis to the field of algorithmic game theory. They are as follows:

(1) We review the recent framework for repeated games with imperfect private monitoring

which has significant applications in markets, such as secret price cutting among firms.

(2) We provide the time complexity of verifying a finite state equilibrium in a repeated game with imperfect private monitoring given a path preautomata of all players and an initial correlation device.

(3) We identify the complexity class of verifying a finite state equilibrium in a repeated game with imperfect private monitoring.

1.2 Structure of Thesis

This document is organized as follows. Chapter 2 covers the background of our approach including the difference between one-shot games and repeated games, basic concepts of Nash equilibrium, partially observable Markov decision process (POMDP), and the grim trigger strategy. Chapter 3 shows the related work that has been done so far in repeated games with perfect monitoring, repeated games with imperfect public monitoring and the complexity class of finding Nash equilibria in different types of games. Chapter 4 describes the finite state equilibrium in repeated games with imperfect private monitoring and continues with a detailed example illustrating how to apply POMDP into a repeated game with imperfect private monitoring by using a game called the prisoners' dilemma. Chapter 5 presents the time complexity of verifying a finite state equilibrium in a repeated game with imperfect private monitoring in detail and its complexity class. Finally, Chapter 6 contians discussion and future work.

Chapter 2

Background

In this chapter, we introduce the background of our approach. We discuss the difference between one-shot games and repeated games, the structure of normal form games and the concepts of Nash equilibrium and partially observable Markov decision process (POMDP).

2.1 One-Shot Games and Repeated Games

In game theory, we can divide games into two categories: one-shot games and repeated games.

In one-shot games, each player makes a decision only once, so he will try his best to gain the greatest profit, while in repeated games, strategies will be very different because each player's action choice is based on his observation of other players' action choices in the previous time step.

When we are discussing games in game theory, our analysis mainly focuses on normal form games [40], which are characterized by:

(1) A set of players who will be making decisions: $i = \{1, 2, ..., k\}$ $(k \geq 2)$

(2) A set of pure strategy profiles: $S = S_1 \times S_2 \times ... \times S_k$ where each $S_i = \{a_1, a_2, ..., a_m\}$ is

	Cooperate	Defect
Cooperate	1,1	-1,2
Defect	2,-1	0,0

Table 2.1: The prisoners' dilemma game

the (finite) set of pure strategies (actions) available to player i. Here player i has m strategies (actions) available

(3) A payoff function $U(\sigma) = (U_1(\sigma), U_2(\sigma), ..., U_k(\sigma))$, where $U_i(\sigma)$ $(i = 1, 2, ..., k)$ represents the payoff (or utility) that player i receives when a strategy profile σ is played

There are several classic games discussed in the game theory: Battle of Sexes; Rock, Paper, Scissors; Matching Pennies etc. However, the prisoners' dilemma game is perhaps the best known and most studied game. It helps us understand what governs the balance between cooperation and competition in business, politics, and social settings [13].

Table 2.1 shows a normal form of the prisoners' dilemma game. In the prisoners' dilemma game, players are non-cooperative, and they make decisions simultaneously. Assume there are two players, i (left side) and j (upper side), and both are rational players. Each player has two actions available in each round of the game: Cooperate (C) or Defect (D). In each round, player i and j are required to choose one element from their strategy spaces in absence of knowledge of the other player's choice. Players obtain payoffs according to the payoff function in Table 2.1 and this payoff table is observed by all players in the game.

If both players cooperate, they both receive reward 1 for cooperating. If player i defects while player j cooperates, then player i receives 2 while player j receives -1. Similarly, if player i cooperates while player j defects, then player i receives -1 while player j receives 2. If both players defect, they both receive nothing.

In a one-shot game, both players have complete knowledge of the payoff table and game is played only once, then there is no further repercussions. The best approach for each player is to find an action which yields the greatest reward. So in the prisoners' dilemma game, both players will choose action D and get 0.

However, when playing a repeated game which for current time step, the outcomes of the previous stage games are observed (fully or partially), players can behave differently. Instead of choosing D, both players may get better returns in a long run by both cooperating. So the better strategy in a repeated game is that players keep updating their strategies based on their observations.

2.2 Nash Equilibrium

The Nash equilibrium is a solution concept of a non-cooperative game involving two or more players. If each player has chosen a strategy and no player can benefit by changing strategies while the other players keep theirs unchanged (all players have no incentive to deviate), then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium [40]. In a formal definition, it is

$$\forall i, x_i \in S_i : \pi_i(x_i^*, x_{-i}^*) \geq \pi_i(x_i, x_{-i}^*) \tag{2.1}$$

Here x_i is a strategy profile of player i, x_{-i} represents the strategy profiles of all other players and (x_i^*, x_{-i}^*) constitutes a Nash equilibrium.

When the inequality above holds strictly (with $>$ instead of $\geq$) for all players and all feasible alternative strategies, then the equilibrium is classified as a strict Nash equilibrium. If instead, for some player, there is exact equality between x_i^* and some other strategy in the set S, then the equilibrium is classified as a weak Nash equilibrium [51].

In the one-shot prisoners' dilemma game shown in Table 2.1, the unique Nash equilibrium is for both players to defect because no player can benefit by switching strategies, given that every other player sticks with the same strategy. If player i choose D, the best response for player j is defection since we have $0 > -1$ and if player j choose D, the best response for player i is defection as well.

2.3 Grim Trigger Strategy

Many different kinds of strategies can be used for discussing repeated games, here we use the grim trigger strategy for our analysis.

In the prisoners' dilemma game, if a player is using the grim trigger strategy then initially, he will cooperate, but as soon as his opponent defects (satisfying the trigger condition), the player using grim trigger will defect for the remainder of the iterated game. Since a single defection by the opponent triggers defection forever, grim trigger is the most strictly unforgiving strategy in the repeated game.

2.4 Partially Observable Markov Decision Process

A sequential decision problem for a fully observable, stochastic environment with a Markovian transition model and additive rewards is called a Markov decision process, or MDP. It consists of a set of states with an initial state s_0, a set ACTIONS(s) of actions in each state, a transition model $P(s'|s,a)$ which stands for the probability of moving to state s' in the next time step given current state s and action a, and a reward function $R(s)$ given current state [46].

In MDP, policy is the solution to a certain problem, and it specify what the agent should do for any possible state that the agent might reach. The quality of a policy is therefore measured by the expected utility of the possible environment histories generated by that policy. An optimal policy is a policy that yields the highest expected utility [46].

With the assumption of full observability in MDP, the agent always knows which state it is in. This, combined with the Markov assumption for the transition model, means that the optimal policy depends only on the current state. However, in partially observable Markov decision process (POMDP), when the environment is partially observable, it has a noisy or partial sensor instead of knowing its state correctly, so it cannot execute the optimal action for that state. Furthermore, the utility of a state s and the optimal action in state s depend not just on state s, but also on how much the agent knows when it is in state s. For these reasons, partially observable MDPs (POMDPs) are usually viewed as much more difficult than ordinary MDPs [46].

The framework of POMDP includes six parameters: S, A, T, Ω, O and R respectively.

S : The set of states

A : The set of actions available to the agent

T : $S \times A \rightarrow \Delta(S)$, The transition function from one state to another, given an action and current state

Ω: The set of possible observations

O : $S \times A \rightarrow \Delta(\Omega)$, The observation function which gives the probability that an action yields a given observation

R : $S \times A \rightarrow R$, The reward function for executing an action in a given state

Since the state is not observable at any given time, the agent represents information about the environment by using belief. The probability of perceiving e, given that action a was performed starting in belief state b, is given by summing over all actual states s' that the agent might reach [46]:

$$\begin{aligned}P(e|a,b) &= \sum_{s'} P(e|a,s',b)P(s'|a,b) \\ &= \sum_{s'} P(e|s')P(s'|a,b) \\ &= \sum_{s'} P(e|s') \sum_{s} P(s'|s,a)b(s).\end{aligned} \tag{2.2}$$

Let us write the probability of reaching b' from b, given action a, as $P(b'|b,a)$. Then that gives us:

$$\begin{aligned}P(b'|b,a) = P(b'|a,b) &= \sum_{e} P(b'|e,a,b)P(e|a,b) \\ &= \sum_{e} P(b'|e,a,b) \sum_{s'} P(e|s') \sum_{s} P(s'|s,a)b(s)\end{aligned} \tag{2.3}$$

When we are solving POMDP, we use the Bellman Equation which is

$$V^h(b^t) = \max_a \{\sum_s b^t(s)R(s,a) + \gamma \sum_o Pr(o|b^t,a)V^{h-1}(B.U(b^t,a,o))\} \tag{2.4}$$

$V^h(b^t)$ is the expected reward of an optimal horizon h policy at time step t, where γ is the discount factor which means we discount future rewards so that they are not as important as the current reward and $B.U(b^t,a,o)$ stands for the belief update based on the current belief, action and observation.

In POMDP, a complete solution maps the entire set of beliefs to a corresponding policy. At first, agent computes the optimal action according to its belief and then performs the action that may alter the state of the environment. After that, agent receives observation from the environment, and the agent updates its belief given action and observation. We do this recursively until it satisfies the termination condition.

Chapter 3

Related Work

Repeated games have a wide range of applications, such as collusion between firms, cooperation among workers, and many military purposes. Repeated games with perfect monitoring and repeated games with imperfect public monitoring have received considerable attention from researchers. There are many studies and papers related to these two models and the key assumption in the existing literature is that players share common information about other players' actions [25].

3.1 Repeated Games with Perfect Monitoring

In repeated games with perfect monitoring, all players observe the action profile chosen at the end of each time step. In other words, the actions of every player are perfectly monitored by all other players and any deviations from the equilibrium path of play can be detected and punished.

Repeated games with perfect monitoring are relatively simple since they do not contain any uncertainty. In the subsequent time step, players make their decisions based on their observations. A strategy profile σ is a subgame-perfect equilibrium of the repeated game if for all histories $h^t \in \mathcal{H}$, the strategy profile $\sigma|_{h^t}$ is a Nash equilibrium of the repeated game [34].

There are several important concepts introduced in repeated games with perfect monitoring. Fudenberg and Maskin [16] introduced the folk theorem. It indicates that in the infinitely repeated version of the game, provided players are sufficiently patient, there is a Nash equilibrium such that both players cooperate on the equilibrium path. Abreu, Pearce, and Stacchette [2] discussed the ideas of a self-generating set of equilibrium payoffs. Mailath, Obara, and Sekiguchi [33] provided a detailed analysis on subgame-perfect equilibria in which player j always plays cooperate in the prisoners' dilemma game. Farrell and Maskin [15] illustrated the renegotiation in repeated games with perfect monitoring.

Suppose that players i and j are playing the prisoners' dilemma game repeatedly at time $t = 0,\ 1,\ 2...$ and that player i's payoff for the entire repeated game is:

$$u_i(\{a^1, a^2, ...\}) = (1-\delta)\sum_{t=0}^{\infty} \delta^t g_i(a_i^t, a_{-i}^t) \tag{3.1}$$

The discount factor $\delta \in [0, 1)$ means that players discount the future. The overall payoffs are multiplied by $(1-\delta)$ to get a per-period average payoff for the game which makes the repeated game payoff comparable to the stage game payoffs [29].

The following example is discussed in Levin's tutorials on Repeated Games with Perfect Monitoring [29].

Proposition 1. *If $\delta \geq 1/2$, then the repeated prisoners' dilemma game has a subgame-perfect equilibrium in which (C, C) is played in every period.*

To proof this proposition, we assume that players are using grim trigger strategies (Figure 3.1). We need to verify that there is no single period where player i can make a profitable deviation.

We suppose up to time period t, D has never been played. Then player i's payoff forward

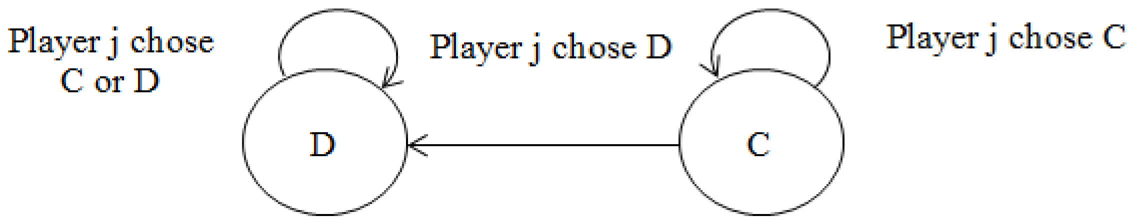

Figure 3.1: Player i is using the grim trigger strategy for the prisoners' dilemma game with perfect monitoring

are:

$$\text{Play C: } (1-\delta)[1+\delta+\delta^2+...] = 1 \tag{3.2}$$

$$\text{Play D: } (1-\delta)[2+\delta\cdot 0+\delta^2\cdot 0+...] = 2(1-\delta) \tag{3.3}$$

We can easily find out if $\delta \geq 1/2$, player i chooses C is optimal.

Suppose that at time period t, D has already been played. Then player j will play D and no matter what will continue to play D, so player i's payoffs are:

$$\text{Play C: } (1-\delta)[-1+\delta\cdot 0+\delta^2\cdot 0+...] = (1-\delta)(-1) \tag{3.4}$$

$$\text{Play D: } (1-\delta)[0+\delta\cdot 0+\delta^2\cdot 0+...] = 0 \tag{3.5}$$

So play D is definitely optimal.

Repeated games with perfect monitoring has been widely used to address questions of economic area. One of the most straightforward applications is the price of a same product in two different places. We consider two supermarkets, A and B. They are in the same neighborhood. They both sell a kind of milk with the same brand, type and size. Supermarket B knows the milk price in Supermarket A, and Supermarket A knows the milk price in Supermarket B. In each time step, they have to decide whether to change the milk price or not. In this case, we can apply a repeated game with perfect monitoring model.

Each supermarket makes a decision on the price of its milk based on the price observed from another supermarket, and each supermarket is trying to make best response to its opponent so that it can gain the greatest profits.

Besides that, Rotemberg and Saloner [44] studied price wars. They established conditions under which oligopolistic firms will behave more competitively when demand is high rather than low. Also, Chari and Kehoe [10] discussed the time inconsistency. Thomas and Worrall [48] and Kocherlakota [27] examined the risk sharing without commitment in perfect monitoring.

3.2 Repeated Games with Imperfect Public Monitoring

In repeated games with imperfect public monitoring, players no longer have the correct information about the past play. The link between current actions and future play is indirect and deviations cannot be unambiguously detected. It is then relatively straightforward to provide incentives for players to not play stage-game best replies [34]. In this class of game, at the end of each time step, all players publicly receive imperfect signals, and in the subsequent time step, they perform their actions based on their current states and public signals observed. We consider the signal to be public because it is observed by all players in the game, and it is imperfect because it contains noisy information.

There are at least two reasons why this information structure is worth special attention [37]. First, it is simply a more reasonable assumption than perfect monitoring in some settings. Second, the equilibrium behavior can be significantly different from the equilibrium behavior in models with perfect monitoring, and they sometimes have significant economic implications.

The general model for repeated games with imperfect public monitoring [30] is shown as follows:

Let $A_1 ..., A_I$ be finite action sets;

Let Y be a finite set of public signals;

Let $\pi(y|a)$ =Pr$(y|a)$;

Let $r_i(a_i, y)$ be i's payoff that he plays a_i and the public signal is y.

Player i's expected payoff is

$$g_i(a) = \sum_{y \in Y} \pi(y|a) r_i(a_i, y) \tag{3.6}$$

A perfect public equilibrium is a profile of public strategies σ that for any public history h^t, specifies a Nash equilibrium for the repeated game, that is, for all t and all h^t, $\sigma|h^t$ is a Nash equilibrium. A perfect public equilibrium is strict if each player strictly prefers his equilibrium strategy to every other public strategy [34].

We again study the repeated prisoners' dilemma game. The following example is studied in Mailath and Samuelson's book on Repeated Games and Reputations [34].

The imperfect monitoring is captured by two signals $\bar{y}$ and $\underline{y}$, whose distribution is given by:

$$Pr(\bar{y}|a) = \begin{cases} p, \text{ if a=CC} \\ q, \text{ if a=CD or DC} \\ r, \text{ if a=DD} \end{cases} \tag{3.7}$$

	$\bar{y}$	$\underline{y}$
Cooperate	$1+\frac{2-2p}{p-q}$	$1-\frac{2p}{p-q}$
Defect	$\frac{2-2r}{q-r}$	$\frac{-2r}{q-r}$

Table 3.1: The payoff table for the prisoners' dilemma game with the public monitoring

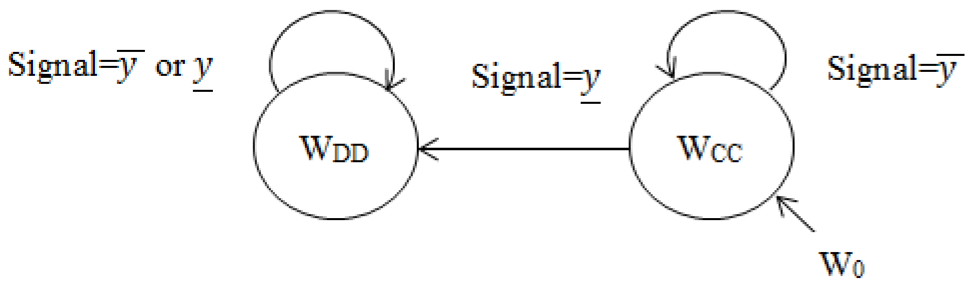

Figure 3.2: The grim trigger strategy for the prisoners' dilemma game with imperfect public monitoring

Since $Pr(\bar{y}|a) + Pr(\underline{y}|a) = 1$, we can also get the $Pr(\underline{y}|a)$ as follows:

$$Pr(\underline{y}|a) = \begin{cases} 1-p, & \text{if a=CC} \\ 1-q, & \text{if a=CD or DC} \\ 1-r, & \text{if a=DD} \end{cases} \tag{3.8}$$

where $0 < q < p < 1$ and $0 < r < p$. $\bar{y}$ has a higher probability when both players cooperate. Payoff table is given by Table 3.1.

We apply the grim trigger strategy again (Figure 3.2), the state space is $\{w_{CC}, w_{DD}\}$ with the initial state w_{CC}. The action choice for each state is $f(w_{CC}) = CC$ and $f(w_{DD}) = DD$.

The transition function is given by

$$\tau(w,y) = \begin{cases} w_{CC}, \text{ if } w = w_{CC} \text{ and } y = \bar{y} \\ w_{DD}, \text{ otherwise} \end{cases} \tag{3.9}$$

So the value function of player i is given by

$$V_i(w_{CC}) = (1-\delta)1 + \delta\{pV(w_{CC}) + (1-p)V(w_{DD})\} \tag{3.10}$$

and

$$V_i(w_{DD}) = (1-\delta)0 + \delta V(w_{DD}) \tag{3.11}$$

We can get $V_i(w_{DD})$=0, and

$$V_i(w_{CC}) = \frac{(1-\delta)}{1-\delta p} \tag{3.12}$$

The grim trigger strategy will be an equilibrium if and only if in each state, the prescribed actions constitute a Nash equilibrium of the normal form game induced by the current payoffs and continuation values. We can have the conditions for equilibrium are:

$$V_i(w_{CC}) \geq (1-\delta)2 + \delta\{qV(w_{CC}) + (1-q)V(w_{DD})\} \tag{3.13}$$

and

$$V_i(w_{DD}) \geq (1-\delta)(-1) + \delta V(w_{DD}) \tag{3.14}$$

It is clear that the incentive constraint for defecting in state w_{DD} is trivially satisfied. The incentive constraint in the state w_{CC} can be rewritten as

$$V_i(w_{CC}) \geq \frac{2(1-\delta)}{1-\delta q} \tag{3.15}$$

We use (3.12) to substitute for $V_i(w_{CC})$, and get

$$\frac{(1-\delta)}{1-\delta p} \geq \frac{2(1-\delta)}{1-\delta q} \tag{3.16}$$

that is,

$$\delta(2p-q) \geq 1 \tag{3.17}$$

so that if $(2p-q) > 1$, then grim trigger is an equilibrium, provided the players are sufficiently patient.

There are some new concepts studied in repeated games with imperfect public monitoring. Abreu, Milgrom and Pearce [1] illustrated the timing of information which shows that with imperfect monitoring, reducing the interest rate always increases the possibilities for cooperation. Mailath et al. [34] [32] discussed private strategies in imperfect public monitoring, that is, strategies that are nontrivial functions of private histories rather than public histories. In addition, folk theorem has also been analyzed in repeated games with imperfect public monitoring [16].

Repeated auctions are important applications of repeated games with imperfect public monitoring. In the first round, each bidder makes his own decision by giving out the price. At the end of this round, all bidders do not know what other bidders' prices. However, Aoyagi [3] considers a model of repeated auction where bidders communicate their bids at the end of each round. In this case, we can apply a repeated game with imperfect public monitoring model.

Since communications may contain fake information, we consider it to be a noisy public signal. In the following time step, each bidder makes his own decision (i.e. either to bid or to give up) based on his current state and imperfect public signal observed.

In addition, Green and Porter [19] studied price wars with public monitoring. They talked about the noncooperative collusion and under imperfect price information. Radner

[42] and Rubinstein [45] studied the applications of repeated games with imperfect public monitoring which involves risk aversion of a firm owner. Levin [28] studied the design of self-enforced relational contracts.

3.3 Complexity Class of Finding Nash Equilibria in Different Types of Games

In recent years, determining the complexity class of finding Nash Equilibria in different types of games has attracted much interest in computer science and economics.

Gottlob et al. [18] indicated that determining whether a strategic game has a pure Nash equilibrium is NP-complete and remains NP-complete even for following two restricted cases: (1) Games in graphical normal form having bounded neighborhood and (2) Acyclic-graph games, and acyclic-hypergraph games.

Borgs et al. [6] proved that, in the general case, computing the Nash equilibria for a repeated game is not easier than computing the Nash equilibria for one-shot finite games, a problem which lies in the PPAD complexity class. Daskalakis, Goldberg and Papadimitriou [12] showed that finding a Nash equilibrium in three-player games is indeed PPAD-complete, and they did so by a reduction from Brouwer's problem, thus establishing that the two problems are computationally equivalent. By improving their construction, they showed even two-player games are also PPAD-complete.

Partially observable stochastic games (POSGs) generalize the repeated games studied in this document to a sequential state that is not perfectly observable. In this class of games, multiagent generalizations of POMDPs [14] are making important contributions. The complexity class of multiagent POMDP frameworks such as the decentralized POMDP (DEC-POMDP) has also been identified. Bernstein et al. [5] considered decentralized control of Markov decision processes and gave complexity bounds on the worst-case running time for algorithms that find optimal solutions. They indicated that for two players in the game, the decision problem of DEC-POMDP is NEXP-complete and it is at least this hard when there are more players.

Chapter 4

Finite State Equilibrium in Repeated Games with Imperfect Private Monitoring

Repeated games with imperfect private monitoring represent long-term relationships. At the end of each time step, each player receives a private signal that indicates what other players' actions were in the current time step. The signal is not fully accurate, so they can be considered a noisy observation of the opponents' actions.

Due to the lack of common information shared by all players, this class of game is difficult to analyze. However, it has broader applications than repeated games with perfect monitoring or imperfect public monitoring. One good example is the collusion under secret price-cutting. Stigler [47] argued that collusion between firms is hard to sustain when other firms cannot observe "cutting" (deviations from collusion) that a firm may offer customers. For example, there are two stores, A and B. They are selling the same product privately. When we apply the framework of repeated games with imperfect private monitoring and focus on store A's decision problem, we have:

State: The price of the product offered by B

Action: Raise the price, reduce the price or keep the price unchanged

Signal: The level of sales

If store A observes a low level of sales, A is not sure whether this is because there is an adverse shock to the demand for its product or B is cheating on the Cartel Agreement [49], so the level of sales can be considered a noisy signal.

When we focus on repeated games with imperfect private monitoring, the arguments of perfect monitoring and imperfect public monitoring do not immediately apply. The construction of equilibria in general imperfect private monitoring games require significantly different techniques [34]. We present the partially observable Markov decision process (POMDP) model in this section, which helps to complete the analysis.

We focus on verifying a finite state equilibrium in a repeated game with imperfect private monitoring. A finite state equilibrium is a (correlated) sequential equilibrium of a repeated game with private monitoring, where each player's behavior on the equilibrium path is given by finite state path preautomata $m_i \equiv (\Theta_i, f_i, T_i)$, $i = 1, ..., k$ and an initial correlation device which is a joint probability distribution of the initial states of the automata, $r \in \Delta(\Theta)$ [26].

4.1 Repeated Games with Imperfect Private Monitoring

Repeated games with imperfect private monitoring are much more complicated than repeated games with perfect monitoring and repeated games with imperfect public monitoring. At the end of each time step, each player only receives a private signal that indicates what other players' actions were in current time step, and the signal is not fully accurate, so they can be considered a noisy observation of the opponents' actions.

Assume there are k players in the game, with finite stage game action set for player

$i \in \{1, ..., k\}$ denoted A_i. At the end of each time step, each player i observes a private signal, denoted ω_i, drawn from a finite set Ω_i. The signal vector $\omega \equiv (\omega_1, ..., \omega_k) \in \Omega \equiv \Omega_i \times ... \times \Omega_k$ occurs with probability $q(\omega|a)$ when the action profile $a \in A \equiv \prod_i A_i$ is chosen [34]. Player i's realized payoff is determined by his own action and signal, denoted $U_i(a_i, \omega_i)$, and his expected payoff is given by $\sum_{\omega\in\Omega} U_i(a_i, \omega_i)q(\omega|a)$ [26].

To illustrate the strategy of each player in a more general way, Osborne and Rubinstein [38] introduced path automaton to represent the strategy a player is using. We are interested in equilibria where each player's behavior on the equilibrium can be described by a finite path automaton. A path automaton without the specification of the initial state is referred as a *path preautomaton*. Path preautomaton contains all states available for player i, denoted by Θ_i, transition function T_i and action choice for each state f_i. Figure 4.1 is a path preautomata example of the grim trigger strategy applied in a repeated game with imperfect private monitoring with two players in the game. In this preautomaton, each circle is one player's state and path arrow represents one transition, labeled by a signal leading to the transition. So in state P, player's action is D and if the player receives a private signal c or d, he will stay in state P for the next time step. If the player is in state R, then the available action is C, and the player will stay in state R if he receives private signal c. However, if the signal is d, then he will move to state P and never return.

4.2 Partially Observable Markov Decision Process in Repeated Games with Imperfect Private Monitoring

To verify that a finite path preautomaton profile is a finite state equilibrium in a repeated game with imperfect private monitoring, our analysis is based on player i's decision problem

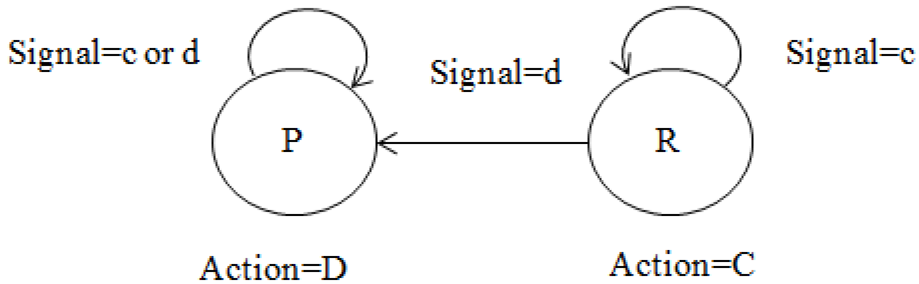

Figure 4.1: This path preautomaton corresponds to the grim trigger strategy. P and R are two automaton states denoting punishment and reward, respectively.

given all opponents' finite state path preautomata m_{-i}. This is complicated by the fact that the opponents' states are not truly observable, and player i only obtains partial information through the private signal ω_i he receives. The POMDP formulation for player i's decision problem contains following:

Θ_{-i} : The set of joint states of all other players

A_i : The set of actions available to player i

T_i : $\prod_{j \neq i} T_i(\theta'_j|\theta_j, w_j)$. The transition function from one state to another based on the current state and a signal received

Ω_i : The set of possible signals received

O_i : $A \to \Delta(\Omega)$. The observation function which gives the probability that a joint action profile yields a given joint signal

R_i : $A \to R_i$. The reward function for executing an action in a given state and it is the expected payoff.

The joint distribution of current signal ω_i and next state θ'_{-i} given the current state θ_{-i}

and action a_i is

$$P(w_i, \theta'_{-i}|\theta_{-i}, a_i) \equiv \sum_{w_{-i}} \prod_{j \neq i} T_i(\theta'_j|\theta_j, w_j) O_i(w_i, w_{-i}|a_i, f_{-i}(\theta_{-i})). \tag{4.1}$$

On the right side of the equation, they are transition function and observation function. Player i's posterior belief about his opponents' states, denoted $P[a_i, \omega_i, b_i](\theta'_{-i})$, can be calculated by his current belief b_i, current action a_i and current signal ω_i. This posterior belief is calculated by

$$\begin{aligned} P[a_i, \omega_i, b_i](\theta'_{-i}) &= \alpha \cdot P(w_i, \theta'_{-i}|b_i, a_i) \\ &= \alpha \cdot \sum_{\theta_{-i}} P(w_i, \theta'_{-i}|\theta_{-i}, a_i) b_i(\theta_{-i}) \end{aligned} \tag{4.2}$$

where α is a normalization constant

$$\alpha = \frac{1}{\sum_{\theta_{-i}} P(w_i|\theta_{-i}, a_i) b_i(\theta_{-i})}. \tag{4.3}$$

4.3 An Example

We implement the POMDP technique to the prisoner's dilemma game introduced in Table 2.1. This example is referred from Kandori and Obara's mimeo [26]. Assume there are two players (i and j), and each player has two states (P and R), two actions (C and D), and in each time step, they may receive either c or d as a private signal. When a player chose action C, the opponent will most likely receive the signal c, but the noisy observation may provide signal d instead. The probability that exactly one player receives an incorrect signal is ε which is greater than 0, and both receive incorrect signals with probability ξ which is also greater than 0. Based on the action profile, we can get the joint distributions of private signals based on different action profiles. We analyze player i's decision problem given player

	c	d
c	$1-2\varepsilon-\xi$	ε
d	ε	ξ

Table 4.1: Joint distribution of private signals based on action profile (C,C)

	c	d
c	ε	$1-2\varepsilon-\xi$
d	ξ	ε

Table 4.2: Joint distribution of private signals based on action profile (D,C)

j's path preautomaton. Player i's path preautomaton m_i corresponds to the grim trigger strategy which is displayed in Figure 4.1.

Let $\varepsilon = 5/36$, $\xi = 1/36$, discount factor $\delta = 0.9$. We can determine the value functions under different joint states by solving the system of linear equations below. The value function of a joint state equals the summation of the immediate reward and the long-term expected reward:

	c	d
c	ε	ξ
d	$1-2\varepsilon-\xi$	ε

Table 4.3: Joint distribution of private signals based on action profile (C,D)

	c	d
c	ξ	ε
d	ε	$1-2\varepsilon-\xi$

Table 4.4: Joint distribution of private signals based on action profile (D,D)

$$\begin{cases} v^{RR} = 1 + \delta\{(1-2\varepsilon-\xi)v^{RR} + \varepsilon v^{RP} + \varepsilon v^{PR} + \xi v^{PP}\} \\ v^{RP} = -1 + \delta\{(1-2\varepsilon-\xi)v^{PP} + \varepsilon v^{RP} + \varepsilon v^{PP} + \xi v^{RP}\} \\ v^{PR} = 2 + \delta\{(1-2\varepsilon-\xi)v^{PP} + \varepsilon v^{PR} + \varepsilon v^{PP} + \xi v^{PR}\} \\ v^{PP} = 0 + \delta\{(1-2\varepsilon-\xi)v^{PP} + \varepsilon v^{PP} + \varepsilon v^{PP} + \xi v^{PP}\} \end{cases} \Rightarrow \begin{cases} v^{RR} = 3.0588 \\ v^{RP} = -1.1765 \\ v^{PR} = 2.3529 \\ v^{PP} = 0 \end{cases}$$

The expected payoff for player i at state θ_i is $V_i^{(m_i,\theta_i)}(b_i)$ where m_i is player i's path preautomaton corresponds to the grim trigger strategy, b_i denotes player i's belief that player j is in state R and $(1-b_i)$ denotes player i's belief that player j is in state P. This linear function can be calculated by

$$V_i^{(m_i,\theta_i)}(b_i) = b_i v^{\theta_i R} + (1-b_i)v^{\theta_i P} \tag{4.4}$$

We can get the following linear functions for player i in state R and P:

$V_i^{(m_i,R)}(b_i) = 3.0588b_i + (1-b_i)\cdot(-1.1764) = 4.2353b - 1.1765$

$V_i^{(m_i,P)}(b_i) = 2.3529b_i + (1-b_i)\cdot 0 = 2.3529b$

The initial value function in the POMDP iteration is the upper envelope of these two linear functions in Figure 4.2 (Red line is $V_i^{(m_i,P)}(b_i)$ and blue line is $V_i^{(m_i,R)}(b_i)$. In the Figure 4.2, X axis represents player i's belief that player j is in state R and Y axis is the corresponding value for each belief. We can see that two lines intersect at $b_i = 0.625$. It means if player i's belief that player j is in state R with $[0, 0.625]$ then he should be in state

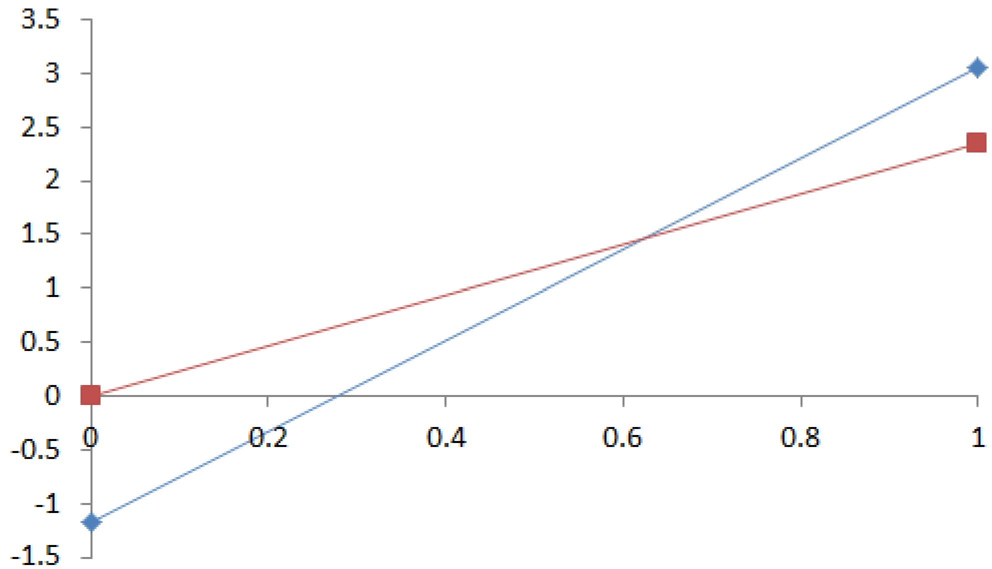

Figure 4.2: The initial value functions of player i

P as the best response. However, if the belief is $[0.625, 1]$ then he should be in state R as the best response.

In the first step of POMDP, we consider the one-shot extensions of player i and see whether they are profitable. A one-shot extension is a one-step deviation from current strategy to another strategy. The reason we consider one-shot extensions is that a strategy profile σ is subgame perfect if and only if there are no profitable one-shot deviations. Let $M^{a_i\theta\theta'}$ be a one-shot extension that player starts with an action a_i and moves to state θ after receiving signal c or moves to state θ' after receiving signal d.

The number of one-shot extensions of player i is calculated by $|A_i| \cdot |\theta_i|^{|\omega_i|} = 2 \cdot 2^2 = 8$ ($|A_i|$ is the number of actions available of player i; $|\theta_i|$ is the number of states in player i's path preautomaton; ω_i is the number of possible signals that player i may receive) and they are : $V_i^{CPR}(b_i)$, $V_i^{DPR}(b_i)$, $V_i^{CRR}(b_i)$, $V_i^{CPP}(b_i)$, $V_i^{DRR}(b_i)$, $V_i^{DRP}(b_i)$, $V_i^{CRP}(b_i)$ and $V_i^{DPP}(b_i)$ respectively. Among them, $V_i^{CRP}(b_i)$ and $V_i^{DPP}(b_i)$ implement the same strategies as $V_i^{C}(b_i)$

and $V_i^D(b_i)$.

Let $P_{a_i\omega_i}(b)$ denote the posterior belief of the opponent on state R when the action, signal and belief of the opponent in state R and P are a_i, ω_i, b and $(1-b)$. Based on Equation 4.1 to Equation 4.3, we can easily get:

$$P_{Cc}(b) = \frac{b \times (1-2\varepsilon-\xi)}{b \times (1-\varepsilon-\xi) + (1-b) \times (\varepsilon+\xi)} \tag{4.5}$$

$$P_{Cd}(b) = \frac{b \times \varepsilon}{b \times (\varepsilon+\xi) + (1-b) \times (1-\varepsilon-\xi)} \tag{4.6}$$

$$P_{Dc}(b) = \frac{b \times \varepsilon}{b \times (1-\varepsilon-\xi) + (1-b) \times (\varepsilon+\xi)} \tag{4.7}$$

$$P_{Dd}(b) = \frac{b \times \xi}{b \times (\varepsilon+\xi) + (1-b) \times (1-\varepsilon-\xi)} \tag{4.8}$$

Let $V_i^{a_i\theta\theta'}(b_i)$ be the expected payoff to player i when he plays his automaton $M^{a_i\theta\theta'}$ under belief b_i. This is a linear function in belief b_i

$$V_i^{a_izz'}(b_i) = b_i v^{a_izz',R} + (1-b_i)v^{a_izz',P} \tag{4.9}$$

where $v^{a_i\theta\theta',R}$ is the payoff under automaton $M^{a_i\theta\theta'}$ when the opponent j plays (m_j, R) and $v^{a_i\theta\theta',P}$ is the payoff under automaton $M^{a_i\theta\theta'}$ when the opponent j plays (m_j, P).

The calculation of the $v^{a_i\theta\theta',R}$ and $v^{a_i\theta\theta',P}$ above is determined by POMDP value function, $V^R(b)$ is the current value function and δ is a discount factor. For example,

$$v^{CRR,R} = R(C,C) + \delta[V^R(P_{Cc}(1))P(\omega_i = c|CC) + V^R(P_{Cd}(1))P(\omega_i = d|CC)] \tag{4.10}$$

$$v^{CRR,P} = R(C,D) + \delta[V^R(P_{Cc}(0))P(\omega_i = c|CD) + V^R(P_{Cd}(0))P(\omega_i = d|CD)] \tag{4.11}$$

We get $v^{CRR,R}$=3.1176, $v^{CRR,P}$=-2.0589, and all other value functions: $v^{CPP,R}$=2.7647, $v^{CPP,P}$=-1, $v^{DRR,R}$=1.5764, $v^{DRR,P}$=-1.0589, $v^{CRP,R}$=1.7059 and $v^{DRP,P}$=-0.17648.

Since $V_i^{CPR}(b_i)$ and $V_i^{DPR}(b_i)$ are dominated by others, here we show the value functions of the remaining six one-shot extensions:

$V_i^{CRR}(b_i) = 5.1765b - 2.0589$

$V_i^{CPP}(b_i) = 3.3647b - 1$

$V_i^{DRR}(b_i) = 2.6353b - 1.0589$

$V_i^{DRP}(b_i) = 1.88238b - 0.17648$

$V_i^{CRP}(b_i) = 4.2353b - 1.1765$

$V_i^{DPP}(b_i) = 2.3529b$

Since player i has two states, we have to solve two systems of linear inequalities and determine the belief space that the given path preautomaton of player i is optimal after the one-shot extension.

$$\begin{cases} 4.2353b - 1.1765 \geq 5.1765b - 2.0589 \\ 4.2353b - 1.1765 \geq 3.3647b - 1 \\ 4.2353b - 1.1765 \geq 2.6353b - 1.0589 \\ 4.2353b - 1.1765 \geq 1.88238b - 0.17648 \\ 4.2353b - 1.1765 \geq 2.3529b \\ b \geq 0 \\ -b \geq -1 \end{cases} \Rightarrow \begin{cases} -0.9412b + 0.8824 & \geq 0 \\ 0.8706b - 0.1765 & \geq 0 \\ 1.6b - 0.1176 & \geq 0 \\ 2.35292b - 1.00002 & \geq 0 \\ 1.8824 - 1.1765b & \geq 0 \\ b & \geq 0 \\ -b + 1 & \geq 0 \end{cases}$$

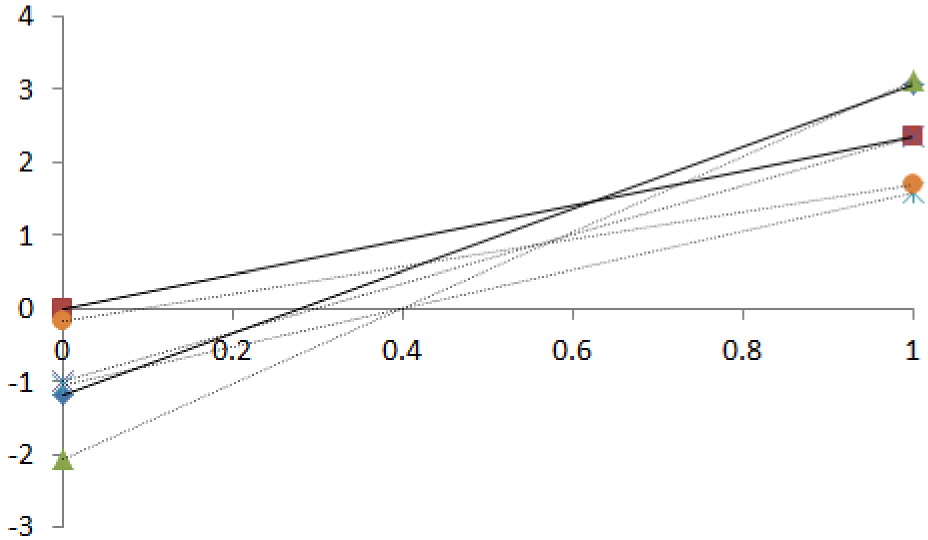

Figure 4.3: The value functions of grim trigger strategy and all other one-shot extensions of player i's path preautomaton

$$\begin{cases} 2.3529b \geq 5.1765b - 2.0589 \\ 2.3529b \geq 3.3647b - 1 \\ 2.3529b \geq 2.6353b - 1.0589 \\ 2.3529b \geq 1.88238b - 0.17648 \\ 2.3529b \geq 4.2353b - 1.1765 \\ b \geq 0 \\ -b \geq -1 \end{cases} \Rightarrow \begin{cases} -2.8236b + 2.0589 & \geq 0 \\ -1.0118b + 1 & \geq 0 \\ -0.2824b + 1.0589 & \geq 0 \\ 0.47052b + 0.17648 & \geq 0 \\ -1.8824b + 1.1765 & \geq 0 \\ b & \geq 0 \\ -b + 1 & \geq 0 \end{cases}$$

As we can see from Figure 4.3, the grim trigger strategy (bold lines) dominates all the other one-shot extensions in the belief region [0, 0.9375].

As the number of states and players increases, so does the number of variables in the systems of linear inequalities. If we have k players and each has $|\theta_j|$ states, then there will be

$(|\theta_j|^{k-1} - 1)$ variables and $(|A_i| \cdot |\theta_i|^{|\omega_i|} - 1 + (|\theta_j|^{k-1} - 1) + 1)$ constraints/linear inequalities in each of $|\theta_i|$ systems of linear inequalities where $(|\theta_j|^{k-1} - 1)$ is the number of constraints that are greater or equal to 0 and last 1 is the number of constraints that should be less than or equal to 1.

Chapter 5

Complexity Results

In this chapter, we start with discussing the time complexity of verifying a finite state equilibrium in a repeated game with imperfect private monitoring. Verifying an equilibrium is hard because we need to check that each player has no incentive to deviate in any possible belief he might have on the past histories. We then provide the complexity class of the whole problem.

5.1 Problem Instance

In a repeated game with imperfect private monitoring with k players involved, we are given:
(1) (m, r) where $m = \{m_1, m_2, ..., m_k\}$ is the path preautomata of all players and r is the initial correlation device
(2) Value functions of player i based on her belief about all other players' automata start states
(3) Value functions of all one-shot extensions of player i's path preautomaton

The decision problem is stated as follows: Can we generate a closed belief region consistent with the initial correlation device r that produces a summation of each value function of player i minus each one-shot extension of player i that is at least zero (in form of $\sum Ab - C \geq 0$, b is the variable, A is the coefficient of variable and C is a constant)? If so, (m, r) is a finite state equilibrium (FSE).

We define a closed belief region such that no profitable one-shot extension exists. The summation of all subtractions to be greater than or equal to zero means that player i's path preautomaton is the best response to other players' path preautomata. Consistency means the initial correlation device r is inside the closed belief region.

Let FSE_k denote the decision problem for the k-agent FSE.

We provide a three-step procedure for an algorithm to verify a FSE:

(1) Find a belief region by solving the following system of linear inequalities, $V_i^{(m_i,\theta_i)}(b_i) \geq V_i^{M_i}(b_i)$, for all M_i in the set of one-shot extensions (This gives us the belief region where no one-shot extensions of player i's path preautomaton are profitable.)

(2) Check the closedness of the belief region which guarantees that after doing two-shot extensions, three-shot extensions and so on, the given path preautomata are still optimal

(3) Check the consistency of the initial correlation device r with the belief region

We also need to perform these steps for all other players.

5.2 The Algorithm

The three step procedure takes exponential time with respect to the number of states of other players' path preautomata in the game. We discuss it in detail.

Step 1: Find a belief region

In the first step of the process, we try to find a belief region which has no profitable one-shot extensions by solving systems of linear inequalities.

Given: The value functions of player i, $V_i^{(m_i,\theta_i)}$; The value functions of all one-shot extensions in the set of all one-shot extensions, Ex(m_i); The number of players k in the game

Question: Do belief regions $B_i^{\theta_i}$ exist for which $V_i^{(m_i,\theta_i)}(b_i) \geq V_i^{M_i}(b_i), \forall M_i \in Ex(m_i)$?

If the answer is YES, then go to the next step and check whether the belief region is a closed belief region. If the answer is NO to any state of player i's preautomaton, then the given path preautomata m cannot constitute a finite state equilibrium because a profitable one-shot extension exists and therefore the preautomaton is not optimal.

Finding the belief region is equivalent to finding the feasible region of the system of linear inequalities. We can find the feasible region by determining the corners of the feasible convex polytope. There are several algorithms that can accomplish this. One efficient and practical way is to introduce slack variables [8]. Slack variables were originally designed for solving optimization problems, but in this case, there is no objective function. We are interested in feasible regions only that satisfy the systems of linear inequalities.

Based on the example in Section 4.3, all constraints are in terms of inequalities, so we first rewrite all inequalities as equations by introducing slack variables s_i, one for each inequality. Since player i has two states R and P, the transformation will be the following respectively:

$$\begin{cases} -0.9412b + 0.8824 & \geq 0 \\ 0.8706b - 0.1765 & \geq 0 \\ 1.6b - 0.1176 & \geq 0 \\ 2.35292b - 1.00002 & \geq 0 \\ 1.8824 - 1.1765b & \geq 0 \\ b & \geq 0 \\ -b + 1 & \geq 0 \end{cases} \Rightarrow \begin{cases} -0.9412b + 0.8824 - s_1 & = 0 \\ 0.8706b - 0.1765 - s_2 & = 0 \\ 1.6b - 0.1176 - s_3 & = 0 \\ 2.35292b - 1.00002 - s_4 & = 0 \\ 1.8824b - 1.1765 - s_5 & = 0 \\ b - s_6 & = 0 \\ -b + 1 - s_7 & = 0 \end{cases}$$

$$\begin{cases} -2.8236b + 2.0589 & \geq 0 \\ -1.0118b + 1 & \geq 0 \\ -0.2824b + 1.0589 & \geq 0 \\ 0.47052b + 0.17648 & \geq 0 \\ -1.8824b + 1.1765 & \geq 0 \\ b & \geq 0 \\ -b + 1 & \geq 0 \end{cases} \Rightarrow \begin{cases} -2.8236b + 2.0589 - s_8 & = 0 \\ -1.0118b + 1 - s_9 & = 0 \\ -0.2824b + 1.0589 - s_{10} & = 0 \\ 0.47052b + 0.17648 - s_{11} & = 0 \\ -1.8824b + 1.1765 - s_{12} & = 0 \\ b - s_{13} & = 0 \\ -b + 1 - s_{14} & = 0 \end{cases}$$

Note that we must have $s_1, s_2 ... s_7$, each greater than or equal to 0. For solving the first system of linear equations, we construct a basic solutions table that shows variable b and slack variables $s_1, s_2 ... s_7$. Set variables (including slack variables) equal 0 and choose n zeros at a time, that is, C_n^{m+n} basic solutions in total where n is the number of variables (excluding slack variables) and m is the number of slack variables which is equal to the number of linear equations in each system of linear equations. The basic solutions table for the first system of linear equations is shown as Table 5.1. Each row represents one calculation of a system of linear equations based on the given variable. For example, in the first row of the basic solutions table, we are given b equals to 0 and have to figure out s_1 to s_7.

We solve each system of linear equations with m variables and m constraints and only those basic solutions with all nonnegative slack variables will be corner points of the feasible region.

The time complexity of solving one system of linear equations with m constraints and m variables is $O(m^3)$ by using Gaussian elimination [17], and checking all slack variables to be nonzero in one basic solution is $O(m)$. So the time complexity for finding belief regions in $|\theta_i|$ systems of linear inequalities requires $O(|\theta_i| \cdot C_n^{m+n} \cdot (m^3 + m)) = O(|\theta_i| \cdot C_n^{m+n} \cdot m^3)$ where C_n^{m+n} is the number of basic solutions we get. Meanwhile, we can generate a polytope

b	s_1	s_2	s_3	s_4	s_5	s_6	s_7
0							
	0						
		0					
			0				
				0			
					0		
						0	
							0

Table 5.1: The basic solutions table for the first system of linear equations

with C_n^{m+n} corner points at most.

In our case, given k players in the game, there will be $n=(|\theta_j|^{k-1}-1)$ variables and $m=(|A_i|\cdot|\theta_i|^{|\omega_i|}-1+|\theta_j|^{k-1})$ constraints (linear inequalities) in one system of linear inequalities. We have $|\theta_i|$ systems of linear inequalities.

The whole time complexity is $f(|\theta_j|)=O(|\theta_i|\cdot C_n^{m+n}\cdot(|A_i|\cdot|\theta_i|^{|\omega_i|}-1+|\theta_j|^{k-1})^3)$. Since C_n^{m+n} is bounded by 2^{m+n}, so this complexity is exponential in the number of states of other players' preautomata.

Step 2: Check the closedness of the belief region

In the second step, we have to check whether the set of beliefs is closed under belief updates.

Proposition 2. *Checking the closedness of the belief region can guarantee that after doing two-shot extensions, three-shot extensions, so on, the given path preautomata remain optimal.*

Mathematically, $V_i^{(m_i,\theta_i)}(b_i) \geq V_i^{M_i'}(b_i), \forall M_i' \in Ex(Ex(m_i))$, $V_i^{(m_i,\theta_i)}(b_i) \geq V_i^{M_i''}(b_i), \forall M_i'' \in Ex(Ex(Ex(m_i)))$, *so on.*

Proof. Let $B_i^{\theta_i}$ be the region of belief for which $V_i^{(m_i,\theta_i)}(b_i) \geq V_i^{M_i}(b_i)$, for all one-shot extensions M_i in $Ex(m_i)$. Let $B_i^{\theta_i}$ be closed under the belief update. Now consider any two-shot extension, M_i', with starting belief in $B_i^{\theta_i}$. After following the action and an observation, the updated belief will still be in $B_i^{\theta_i}$ because the region is closed, and we will arrive at a one-shot extension in $Ex(m_i)$. However, because the controller m_i is better than any one-shot extension in the belief region, we can replace the one-shot extended controller with the original controller and get a better quality two-shot extension. But this two-shot extension now becomes just a one-shot extension to the original controller, m_i. As we previously showed that m_i is better than any one-shot extension on $B_i^{\theta_i}$, it will remain better than the previously constructed one-shot extension as well. This argument can be extended inductively to any n-shot extensions.

In the above argument, it is important that $B_i^{\theta_i}$ be closed. If it is not, then a two-shot extension $M_i'(b_i)$ where b_i is in $B_i^{\theta_i}$ after the action and an observation may lead to a belief that is not in $B_i^{\theta_i}$. Then the original controller, m_i, may not be better than any M_i for the updated belief. □

Given: Prior belief regions $B_i^{\theta_i}$ from Step 1

Question: Is each posterior belief still inside the prior belief region?

If the answer is YES, then the prior belief region is a closed belief region. We can proceed to the third step for checking the consistency of initial correlation device r. If the answer is NO, then the belief region from the first step is not a closed belief region. In other words, after belief updates, the posterior belief will go outside the initial belief region meaning that there exist profitable n-shot extensions. Therefore, player i's path preautomaton m_i is not optimal, so m cannot constitute a finite state equilibrium. To check for closedness, we use the following lemma.

Lemma 1. [Phelan and Skyrzpacz [41]] All posterior beliefs are a convex combination of posterior beliefs of all extreme priors.

For any history sequence, if prior belief b_i is a convex combination of priors b_i' and b_i'', then after applying the belief updates, the posterior of b_i is a convex combination of the posteriors of b_i' and b_i''. Therefore, in our case, we can focus solely on extreme priors and determine whether each extreme prior is still in the initial belief region after the belief update.

We get the time complexity of checking the closedness of a belief region for more than two players and each player has more than two states starting from the following three basic cases:

(1) Two players in the game and the opponent has two states in its preautomaton (two-dimensional belief space)

(2) Two players in the game and the opponent has three states in its preautomaton (three-dimensional belief space)

(3) Two players in the game and the opponent has more than three states in its preautomaton (higher-dimensional belief space)

Case 1: There are two players in the game, player j has two states in its preautomaton (i.e. $|\theta_j|$=2, two-dimensional belief space).

The overall belief region and simplex for this case is shown in Figure 5.1 (bold and dotted line). b_1 represents player i's belief that his opponent is in one state, and b_2 represents player i's belief that his opponent is in another state. The summation of these two beliefs is 1.

Figure 5.1 (bold line) shows a possible belief region after solving the system of linear inequlities in Step 1. It has n corners (n=2). The time complexity of checking whether the posterior beliefs are in the initial belief region is $O(|\theta_j| \cdot n \cdot |\Omega_j| \cdot |\theta_j| + n \cdot n)$ where the first $|\theta_j| \cdot n$ is the number of belief updates we have to perform. $|\Omega_j| \cdot |\theta_j|$ is the time complexity of

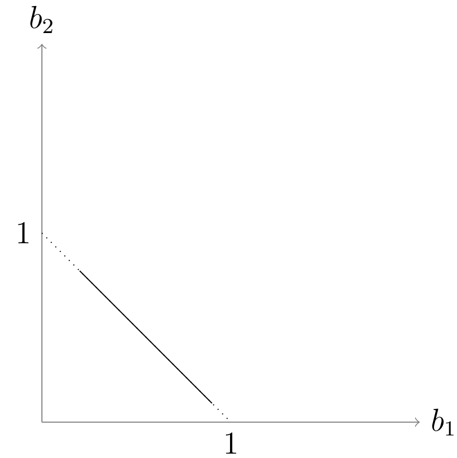

Figure 5.1: An example of a belief region of two states

doing one belief update. The second n is the number of corner points we have, and the third n is the time complexity of checking whether a posterior belief is in the initial belief region or not by simply calculating the summation of the distance with two corners. If it equals to the distance between two corner points, then it is on the line, otherwise, it is not. The whole time complexity is polynomial in the number of states of other players' preautomata.

Case 2: There are two players in the game, player j has three states in its preautomaton.(i.e. $|\theta_j|$=3, three-dimensional belief space).

In this case, the overall belief space is shown in Figure 5.2 (dotted line). b_1 represents player i's belief that his opponent is in the first state, b_2 represents player i's belief that his opponent is in the second state. b_3 represents player i's belief that his opponent is in the third state. The summation of these three beliefs is 1.

Given a initial belief region (convex polygon) with n corners, for example, in Figure 5.2 (bold polygon), the complexity of checking whether posterior beliefs are in the initial belief region is $O(|\theta_j| \cdot n \cdot |\Omega_j| \cdot |\theta_j| + n \cdot n)$ where the first $|\theta_j| \cdot n$ is the number of belief updates we have to perform and $|\Omega_j| \cdot |\theta_j|$ is the time complexity of doing one belief update. The

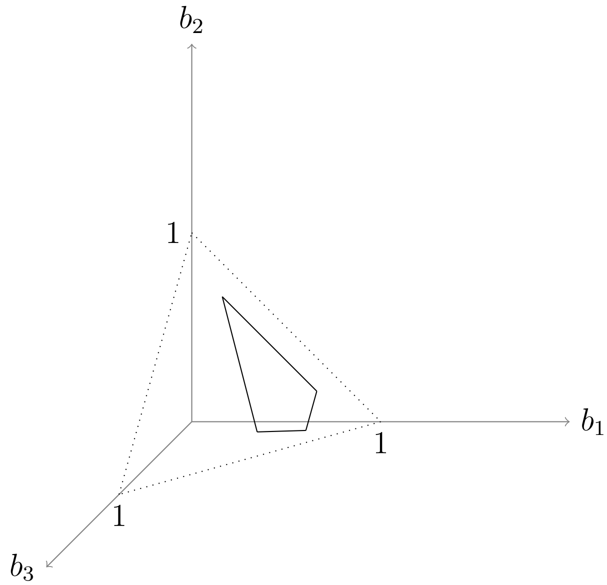

Figure 5.2: An example of a belief region of three states

second n is the number of corners in the plane, and the third n comes from the algorithm for determining if a point P is inside a convex polygon in a three-dimensional space.

To determine whether a point P is in the interior of a convex polygon in three-dimensional space [7], we should first determine whether the point is on the plane, then determine its interior status. Both of these can be accomplished by computing the sum of the angles between the test point and every pair of corner points. If the sum is 2π, then the point is on the plane of the polygon and on the interior (Figure 5.3). The angle sum will tend to be 0 the further point P is away from the polygon. The whole time complexity is polynomial on the number of states of other players' preautomata.

Case 3: There are two players in the game, player j has more than three states in its preautomaton. (i.e. $|\theta_j| > 3$, higher-dimensional belief space).

Given an initial belief region (convex polyhedron) with $|\theta_j|$ dimensions and n corners, the complexity of checking whether the posterior beliefs are in the initial belief regions is

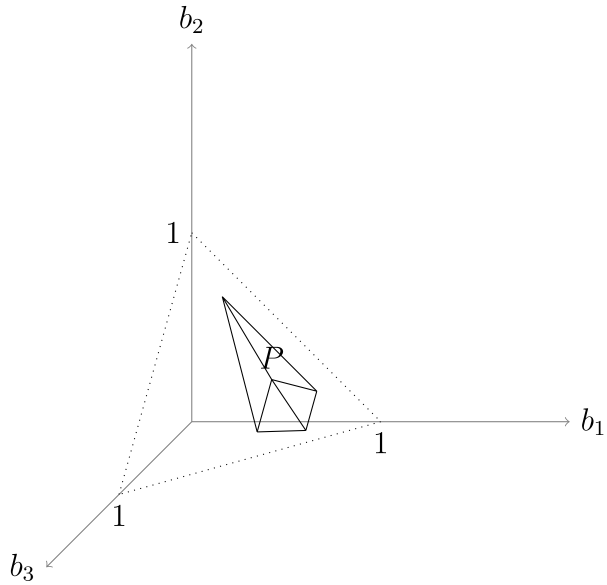

Figure 5.3: The sum of angles is 2π if the point P is in the plane of the 3D polygon and on the interior

$O(|\theta_j| \cdot n \cdot |\Omega_j| \cdot |\theta_j| + n \cdot (|\theta_j| \cdot n))$ where $|\theta_j| \cdot n$ is the number of belief updates we have to perform and $|\Omega_j| \cdot |\theta_j|$ is the time complexity of doing one belief update. We can use a triangle algorithm to determine whether a point is in the polytope with the time bounded by $O(|\theta_j| \cdot n)$ [24, 31]. The whole time complexity is polynomial in the number of states of other players' preautomata.

The triangle algorithm [24, 31] shown in Algorithm 1 is designed to solve the following convex hull decision problem: Given a set of points S $= (v_1, ..., v_n) \subset \mathbb{R}^m$ (m is the number of dimensions) and a distinguished point p $\in \mathbb{R}^m$, test if p lies in $conv(S)$. In each iteration of the triangle algorithm we have a current approximation $p' \in conv(S)$. Using this, we select a pivot $v_j \in S$, i.e. $d(p', v_j) \geq d(p, v_j)$, where $d(.,.)$ is the Euclidean distance.

In summary, for two players, the time complexity of checking whether all posterior beliefs

Algorithm 1 Triangle Algorithms

Input: S=$\{v_1, ..., v_n\}$, a point p and a tolerance $\epsilon \in (0, 1)$

Output: α, the convex combination of $\{v_1, ..., v_n\}$ for p if p $\in conv(S)$ or No, if p $\notin conv(S)$

R=max $d(p, v_j)$ $j \in \{1, ...n\}$, k=0;

Arbitrarily choose initial point $p_0 \in conv(S)$ and corresponding α_0;

while $(d(p, p_k) > \epsilon R)$ **do**

    find v_j such that $d(p, v_j) \leq d(p_k, v_j)$;

    if no such v_j **then**

        return NO and halt;

    else

        compute p_{k+1}, the projection of p on line segment $p_k v_j$, update α_{k+1};

        k=k+1;

    end if

end while

Output α_k;

are in the initial belief region is

$$f(|\theta_j|) = \begin{cases} O(|\theta_i| \cdot (|\theta_j|^2 \cdot n \cdot |\Omega_j| + n^2)) & \text{if } |\theta_j| = 2 \text{ or } 3 \\ O(|\theta_i| \cdot (|\theta_j|^2 \cdot n \cdot |\Omega_j| + |\theta_j| \cdot n^2)) & \text{if } |\theta_j| > 3 \end{cases} \tag{5.1}$$

where n is the number of corners in the polytope, $|\theta_j|$ is the number of states of player j's path preautomaton as well as the number of dimensions of the belief space. $|\theta_i|$ is the number of states of player i's path preautomaton. $|\Omega_j|$ is the size of set of signals of player j and we have $|\theta_i|$ belief regions.

For more than two players, let k be the number of players and each player has at least two states, the time complexity is $f(|\theta_j|) = O(|\theta_i| \cdot [|\theta_j|^{k-1} \cdot n \cdot [|\Omega_j|^{k-1} \cdot (k-1) \cdot |\theta_j|^{k-1}] + n \cdot (|\theta_j|^{k-1} \cdot n)])$, where the first $|\theta_j|^{k-1}$ is for each corner, the number of belief updates we have to perform, n is the number of corners, $[|\Omega_j|^{k-1} \cdot (k-1) \cdot |\theta_j|^{k-1}]$ is the complexity of doing one belief update, $(|\theta_j|^{k-1} \cdot n)$ is the complexity of checking whether a posterior belief is in the initial belief region or not and $|\theta_i|$ is the number of regions we have to check. We can see that the time complexity is polynomial in the number of states of other players' preautomata.

Step 3: Check the consistency of the initial correlation device

In the third step, we check the consistency of initial correlation device r.

Given: Closed belief regions $B_i^{\theta_i}$ and initial correlation device r

Question: Is the correlation device r inside the closed belief regions?

If the answer is YES, then the given (m, r) is a finite state equilibrium, but if the answer is NO, then (m, r) is not a finite state equilibrium.

A correlation device $r \in \Delta(\prod_{i \in k} \Theta_i)$ is a joint probability distribution, which chooses $\theta = (\theta_1, \ldots, \theta_k) \in \prod_{i \in k} \Theta_i$ with probability $r(\theta)$, and recommends θ_i for player i as i's initial state in its preautomaton. We can calculate i's belief for the initial joint states of other players given player i's state by using the Bayes' rule with time complexity $O(|\theta_j|^{k-1} \cdot |\theta_i|)$.

For example, let the initial correlation device be given as r(R,R)= 3/5; r(P,P) =1/5; r(R,P) = r(P,R) = 1/10, when player 1 is recommended R, he knows player 2 is recommended R (or P) with probability

$$P(R_2|R_1) = \frac{P(R_1R_2)}{P(R_1)} = (3/5)/(7/10) = \frac{6}{7} \tag{5.2}$$

$$P(P_2|R_1) = 1 - P(R_2|R_1) = \frac{1}{7} \tag{5.3}$$

Also, when player 1 is recommended P, he knows player 2 is recommended R (or P) with probability

$$P(R_2|P_1) = \frac{P(P_1R_2)}{P(P_1)} = (1/10)/(3/10) = \frac{1}{3} \tag{5.4}$$

$$P(P_2|P_1) = 1 - P(R_2|P_1) = \frac{2}{3} \tag{5.5}$$

Checking whether the initial correlation device is in the closed belief region is analogous to verifying whether a point is in the polytope. So the time complexity is $O(|\theta_i| \cdot (|\theta_j|^{k-1} \cdot n))$ where $|\theta_i|$ is the number of points to check and if all of them are in the corresponding closed belief regions, then (m, r) is a finite state equilibrium. In all, the time complexity is $f(|\theta_j|) = O(|\theta_j|^{k-1} \cdot |\theta_i| + |\theta_i| \cdot (|\theta_j|^{k-1} \cdot n))$ which is polynomial in the number of states of other players' preautomata.

The whole algorithm is shown in Algorithm 2. All parameters are defined as follows:

k: the number of players in the game

n: the number of corners of the polytope

$|\theta_j|$: size of states of player j's preautomaton

In all, the time complexity is not polynomial: it takes exponential time in the number of states of other players' preautomata.

Algorithm 2 An algorithm for verifying finite state equilibrium in the repeated games with imperfect private monitoring

Given the value function of player i, $V_i^{(m_i,\theta_i^l)}$; All value functions of one-shot extensions Ex(m_i); Initial correlation device r; The number of players k
Use the method in Step 1 to determine the corner points of $|\theta_i|$ belief regions;
if there are no corner points exist **then**
    return m can not constitute a finite state equilibrium;
else
    Use the method in Step 2 to determine whether belief regions are closed;
    if they are closed belief regions **then**
        Use the method in Step 3 to check for the consistency of initial correlation device r;
        if r is consistent **then**
            return (m, r) is a finite state equilibrium;
        else
            return (m, r) is not a finite state equilibrium;
        end if
    else
        return Some or all belief regions are not closed, and m cannot constitute a finite state equilibrium;
    end if
end if

5.3 Analysis of Complexity Class

As we saw, given (m, r) and the number of players in the game, it is hard to find a closed belief region that is consistent with the initial correlation device r: The algorithm we provide requires exponential time complexity for verifying that a given (m, r) is a finite state equilibrium.

Theorem 1. *For given k players in a game, FSE_k is in NP if P $\neq$ NP.*

Proof. To show that FSE_k belongs to NP, we show that a certificate consisting of a closed belief region can be verified in polynomial time. The verification procedure can be reduced in three steps:

(1) Verify the closedness of the given closed belief region

(2) Verify the consistency of the initial correlation device

(3) Verify there are no profitable one-shot extensions

The first step of the verification algorithm is performing belief updates for all corners of the given closed belief region. We have to check whether all corners are still in the given closed belief region after the belief updates. Since we know the number of corners in the polytope and are given that there are k players in the game, the time complexity is $f(|\theta_j|) = O(|\theta_j|^{k-1} \cdot n \cdot [|\Omega_j|^{k-1} \cdot (k-1) \cdot |\theta_j|^{k-1}] + n \cdot (|\theta_j|^{k-1} \cdot n))$, where the first $|\theta_j|^{k-1}$ is for each corner, the number of belief updates we have to perform, n equal to the number of corners, $[|\Omega_j|^{k-1} \cdot (k-1) \cdot |\theta_j|^{k-1}]$ is the complexity of doing one belief update, and $(|\theta_j|^{k-1} \cdot n)$ is the complexity of checking whether a posterior belief is in the initial belief region or not. The total time complexity is polynomial in the number of states of other players' preautomata.

In the second step, we check the given initial correlation device and see whether it is in the given closed belief region or not. The time complexity is $f(|\theta_j|) = O(|\theta_j|^{k-1} \cdot |\theta_i| + |\theta_i| \cdot (|\theta_j|^{k-1} \cdot n))$ where the first part is the time complexity of applying Bayes' rule for the initial correlation device r, and the second part is the time complexity of checking whether all distributions over joint states of other players for each state of player i are in the given closed belief reigon. This time complexity is polynomial in the number of states of other players' preautomata.

The third step of the verification algorithm involves calculating i's beliefs over other players' joint states given player i's state from r and then evaluating the finite state equilibrium by replacing the variables in each inequality of the systems of linear inequalities (i.e. Is $Ab - C \geq 0$?) with the beliefs. The time complexity is $O(|\theta_i| \cdot (|\theta_j|^{k-1} - 1) \cdot [|A_i| \cdot |\theta_i|^{|\omega_i|} - 1 + |\theta_j|^{k-1}])$ where $(|\theta_j|^{k-1} - 1)$ is the number of variables in each inequality and $(|A_i| \cdot |\theta_i|^{|\omega_i|} - 1 + |\theta_j|^{k-1})$ is the number of constraints as well as the number of linear inequalities in one system and we have $|\theta_i|$ systems of linear inequalities. This step can also be done in polynomial time.

We conclude that verifying if the given closed belief region is closed and consistent with the initial correlation device, and that there are no one-shot extensions, then the algorithm has verified that (m, r) is a FSE. All three steps can be done in polynomial time. Thus, FSE_k is in NP if P $\neq$ NP. □

Observation 1. *Whether FSE_k is NP-Complete remains an open question.*

In order to show whether FSE_k is NP-complete, we first reduce the problem into two subproblems:

(1) Determining the feasibility of the system of linear inequalities

(2) Given a feasible system of linear inequalities, generate all vertices in the polytope of the feasible region.

Determining the feasibility of a system of linear inequalities is NP-hard only for integer programming [39]. In our case, we have rational coefficients and variables. It can be solved in polynomial time by using the ellipsoid algorithm [43].

The second question falls into the general problem of vertex enumeration [23]. The problem of vertex enumeration is defined as follows:

Input: A system of linear inequalities

Output: All corners of the feasible polytope P

For fixed dimensions which we know the dimensions of the polytope P beforehand, Chazelle [11] found a polynomial time algorithm with complexity of $O(m^{\lfloor d/2 \rfloor})$ (m is the number of inequalities in the input and d is the dimensions of the polytope) and it is optimized by the Upper bound Theorem introduced by McMullen [35]. However, for general cases, there exists a reverse search method introduced by Avis and Fukuda [4] which solves the problem for simple polyhedra in polynomial total time. Bremner, Fukuda and Marzetta [9] solve the problem of simplicial polytopes using a polynomial algorithm. For other cases, no NP-Complete problem that has been found has a polynomial time reduction [23]. In our case, since we do not know the dimensions of the belief space beforehand and the belief space is a subset of the simplicial polytope, so our problem remains open.

We conclude that FSE_k is in NP if P $\neq$NP, but whether it is NP-Complete remains open and needs further study.

Chapter 6

Discussion and Future Work

In the thesis, we use one of the most famous and classic games in the game theory: The prisoners' dilemma to illustrate several important concepts, propositions and theorems. We take a close look at three kinds of repeated games in the game theory. Among them, repeated games with perfect monitoring and repeated games with imperfect public monitoring have received considerable study. We analyze repeated games with imperfect private monitoring, which is a new framework and has significant applications in markets.

We use the POMDP to verify a finite state equilibrium in a repeated game with imperfect private monitoring given the path preautomata of all players and an initial correlation device. The fact that the time complexity is not polynomial though the number of players is given is important.

Given the number of players in the game, we conclude that the complexity class of verifying a finite state equilibrium in a repeated game with imperfect private monitoring is in NP if P $\neq NP$, but it is uncertain whether it is NP-Complete.

There are some alternate ways of proving the complexity analysis, but they are not efficient. For example, we can use Fourier-Motzkin elimination [50] for answering the existence of the belief region. It is a mathematical algorithm for eliminating variables from a system of linear inequalities.

Consider a system S of n inequalities with r variables x_1 to x_r, with x_r the variable to be eliminated. The linear inequalities in the system can be grouped into three classes depending on the sign (positive, negative or null) of the coefficient for x_r:

1) Those inequalities that are of the form $x_r \geq b_i - \sum_{k=1}^{r-1} a_{ik}x_k$; denote these by $x_r \geq A_j(x_1...x_{r-1})$, for j ranging from 1 to n_A where n_A is the number of such inequalities.

2) Those inequalities that are of the form $x_r \leq b_i - \sum_{k=1}^{r-1} a_{ik}x_k$; denote these by $x_r \leq B_j(x_1...x_{r-1})$, for j ranging from 1 to n_B where n_B is the number of such inequalities.

3) Those inequalities in which x_r plays no role, grouped into a single conjunction $\emptyset$

The original system is thus equivalent to:

$$max(A_1(x_1...x_{r-1}), ...A_{n_A}(x_1...x_{r-1})) \leq x_r \leq min(B_1(x_1...x_{r-1}), ...B_{n_B}(x_1...x_{r-1})) \wedge \emptyset \quad (6.1)$$

Elimination consists in producing a system equivalent to $\exists x_r S$. Obviously, this formula is equivalent to:

$$max(A_1(x_1...x_{r-1}), ...A_{n_A}(x_1...x_{r-1})) \leq min(B_1(x_1...x_{r-1}), ...B_{n_B}(x_1...x_{r-1})) \wedge \emptyset \quad (6.2)$$

This inequality is equivalent to $n_A n_B$ inequalities $A_i(x_1...x_{r-1}) \leq B_j(x_1...x_{r-1})$, for $1 \leq i \leq n_A$ and $1 \leq j \leq n_B$

We have therefore transformed the original system into another system where x_r is eliminated. Note that the output system has $(n - n_a - n_b) + n_a n_b$ inequalities. In the worst case, the number of ouput inequalities is $n^2/4$ (if $n_a = n_b = n/2$). We repeat the process, and the system of linear inequalities is satisfiable if and only if the maximum of all q_i in inequalities $x_n \geq q_i$ (Assume x_n is the only variable left in the end) is less than the minimum of all q_j in the inequalities $x_n \leq q_j$, and every q_k in the inequalities $0 \geq q_k$ is non-postive. The time complexity of doing these whole process is $O(n^{2^r})$.

Method	Computing Complexity
Laplace expansion	$n!$
Gaussian elimination	$n^3/3 + n^2/2$
Gauss-Jordan	$n^3/3 + n^2 - n/3$
LU decomposition	$n^3/3 + n^2 - n/3$
QR decomposition	$2n^3 + 3n^3$
Single value decomposition	$2n^3 + 4n^3$

Table 6.1: A comparison in direct and iterative methods (algorithms) for solving linear equation systems

So in our case, for k players, there are $(|\theta_j|^{k-1} - 1)$ variables and $(|A_i| \cdot |\theta_i|^{|\omega_i|} - 1 + (|\theta_j|^{k-1} - 1) + 1)$ constraints/linear inequalities in a system of linear inequalities. The time complexity is $f(|\theta_j|) = O([|\theta_j|^{(k-1)}]^{2^{|\theta_j|^k}})$, a doubly exponential complexity, which is very time consuming.

In addition, there are direct and iterative method for solving linear equation systems [21], but most of them have higher time complexity than Gaussian elimination. The comparison is shown in Table 6.1.

Future work will involve designing an algorithm for finding the finite state equilibrium in repeated games with imperfect private monitoring.

Besides that, the complexity class of vertex enumeration with a given system of linear inequalities needs further investigation. So far, there are many algorithms known for the solving the problem of vertex enumeration, but none of them runs in polynomial time for general polytopes.

Bibliography

[1] D. Abreu, P. Milgrom, and D. Pearce. Information and timing in repeated partnerships. *Econometrica*, 59, 1991.

[2] D. Abreu, D. Pearce, and E. Stacchetti. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–269, 1986.

[3] Masaki Aoyagi. Efficient collusion in repeated auctions with communication. *Journal of Economic Theory*, 2002.

[4] David Avis and Komei Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, 8(1):295–313, 1992.

[5] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.

[6] Christian Borgs, Jennifer Chayes, Nicole Immorlica, Adam Tauman Kalai, Vahab Mirrokni, and Christos Papadimitriou. The myth of the folk theorem. 2007.

[7] Paul Bourke. Polygons and meshes. `http://paulbourke.net/geometry/polygonmesh/`.

[8] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[9] David Bremner, Komei Fukuda, and Ambros Marzetta. Primal-dual methods for vertex and facet enumeration. *Discrete & Computational Geometry*, 20(3):333–357, 1998.

[10] V. Chari and P. Kehoe. Sustainable plans and mutual default. *Review of Economic Studies*, 60(1), 1993.

[11] Bernard Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete & Computational Geometry*, 10(1):377–409, 1993.

[12] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[13] Avinash Dixit and Barry Nalebuff. The concise encyclopedia of economics. `http://www.econlib.org/library/Enc/PrisonersDilemma.html`.

[14] Prashant Doshi. Decision making in complex multiagent contexts: A tale of two frameworks. *AI Magazine*, 33(4):82–95, 2012.

[15] J. Farrell and E. Maskin. Renegotiation in repeated games. *Games and Economic Behavior*, 1(4), 1989.

[16] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[17] James E. Gentle. *Numerical Linear Algebra for Applications in Statistics*. Springer, 1998.

[18] Georg Gottlob, Gianluigi Greco, and Francesco Scarcello. Pure nash equilibria: Hard and easy games. *Journal of Artificial Intelligence Research*, 24:357–406, 2005.

[19] E. Green and R. Porter. Noncooperative collusion and under imperfect price information. *Econometrica 52*, 1984.

[20] Edward J. Green and Robert H. Porter. Noncooperative collusion under imperfect price formation. *Econometrica*, 52(1):87–100, 1984.

[21] Stefan Jahn. A comparison in direct and iterative methods (algorithms) for solving linear equation systems. `http://qucs.sourceforge.net/tech/node99.html`.

[22] L.P. Kaelbling, M.L. Littman, and A.R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence Journal*, 101, 1998.

[23] V. Kaibel and M. Pfetsch. Some algorithmic problems in polytope theory. *Algebra, Geometry and Software Systems*, pages 23–47, 2003.

[24] Bahman Kalantari. A characterization theorem and an algorithm for a convex hull problem. *Annals of Operations Research*, pages 1–49, 2012.

[25] Michihiro Kandori. Introduction to repeated games with private monitoring. *Journal of Economic Theory*, 102:1–15, 2002.

[26] Michihiro Kandori and Ichiro Obara. Towards a belief-based theory of repeated games with private monitoring: An application of POMDP. 2010.

[27] N. Kocherlakota. Implications of efficient risk sharing without commitment. *Review of Economic Studies*, 63, 1996.

[28] J. Levin. Relational contracts. *American Economic Review*, 93, 2003.

[29] Jonathan Levin. Repeated games I: Perfect monitoring. 2006.

[30] Jonathan Levin. Repeated games II: Imperfect public monitoring. 2006.

[31] Meng Li and Bahman Kalantari. Experimental study of the convex hull decision problem via a new geometric algorithm.

[32] G. Mailath, S. Matthews, and T. Sekiguchi. Private strategies in finitely repeated games with imperfect public monitoring. *Contributions to Theoretical Economics*, 2(1), 2002.

[33] G. J. Mailath, I. Obara, and T. Sekiguchi. The maximum efficient equilibrium payoff in the repeated prisoners' dilemma. games and economic behavior. *Games and Economic Behavior*, 40(1):99–122, 2002.

[34] George J. Mailath and Larry Samuelson. *Repeated Games and Reputations*. University Press, Oxford, 2006.

[35] Peter McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17(02):179–184, 1970.

[36] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[37] Ichiro Obara. Repeated games with imperfect public monitoring. 2003.

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

[39] Christos Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications, 1982.

[40] Marcello Pelillo and Andrea Torsello. Game theory in computer vision and pattern recognition. `http://www.dsi.unive.it/~atorsell/cvpr2011tutorial/`.

[41] Christopher Phelan. Beliefs and private monitoring. *Review of economic Studies*, 79:1637–1660, 2012.

[42] R. Radner. Repeated principal-agent games with discounting. *Econometrica*, 53(5):1173–1198, 1985.

[43] Steffen Rebennack. Ellipsoid method. *Encyclopedia of Optimization*, pages 890–899, 2008.

[44] J. Rotemberg and G. Saloner. A supergame-theoretic model of price wars during booms. *American Economic Review*, 76(3):390–407, 1986.

[45] A. Rubinstein. An optimal conviction policy for offenses that may have been committed by accident. *Applied Game Theory*, pages 406–413, 1979.

[46] Stuart J. Russell and Peter Norvig. *Artificial Intelligence, A Modern Approach (3rd Edition)*. Pearson Education, Inc., 2010.

[47] G. J. Stigler. A theory of oligopoly. *Journal of Political Economy*, 72(1):4461, 1964.

[48] J. Thomas and T. Worrall. Self-enforcing wage contracts. *Review of Economic Studies*, 55(1), 1988.

[49] Xavier Vives. *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press, 2001.

[50] Wikipedia. Fourier-Motzkin elimination. `http://en.wikipedia.org/wiki/Fourier%E2%80%93Motzkin_elimination`.

[51] Wikipedia. Nash equilibrium. `http://en.wikipedia.org/wiki/Nash_equilibrium`.