Applied Mathematics and Computation 269 (2015) 863–884

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Computing equilibria in discounted dynamic games

Andriy Burkov, Brahim Chaib-draa*

Computer Science Department, Laval University, PQ, Canada

ARTICLE INFO

Keywords:
Game theory
Repeated games
Stochastic games
Markov chain games
Subgame-perfect equilibria
Automaton

ABSTRACT

Game theory (GT) is an essential formal tool for interacting entities; however computing equilibria in GT is a hard problem. When the same game can be played repeatedly over time, the problem becomes even more complicated. The existence of multiple game states makes the problem of computing equilibria in such games extremely difficult. In this paper, we approach this problem by first proposing a method to compute a nonempty subset of approximate (up to any precision) subgame-perfect equilibria in repeated games. We then demonstrate how to extend this method to approximate *all* subgame-perfect equilibria in a repeated game, and also to solve more complex games, such as Markov chain games and stochastic games. We observe that in stochastic games, our algorithm requires additional strong assumptions to become tractable, while in repeated and Markov chain games it allows approximating *all* subgame-perfect equilibria reasonably fast and under considerably weaker assumptions than previous methods.

1. Introduction

Repeated interactions are studied in Biology, Economics, Ecology, Social Sciences, Computer Science and many other domains. The theory of repeated games allows us to model such interactions by considering a group of agents (or players – both are interchangeable in this paper) evolving in a strategic interaction over and over. Notice that in repeated games with a *complete information*, the data of the strategic interaction is fixed over time and is known by all the players (this is the case of this paper). Usually, to solve a multi-player game means finding a particular strategy for each player, such that the collection of players' strategies forms an *equilibrium*. Just like in a single-agent case, each agent's strategy has to be preferred by that player over any other strategy, assuming the other players' strategies and the environment characteristics remain constant.

Dynamic games, such as repeated games [1,2], or stochastic games [3,4], involve multiple stages of interactions between players. However when these games are played by players having bounded rationality and do not share all the other severe assumptions sustained by the classical theory game, it is evolutionary game theory (EGT) [2,4,5] which used as theory of dynamic adaptation and learning. EGT is generally used for evolving populations (as for instance in [6]) and in this sense it focuses more on the dynamics of strategy change as influenced not solely by the quality induced by competing strategies, but also by the frequency with which those strategies are found in the population [7]. EGT uses an equilibrium refinement of the Nash equilibrium, called *evolutionarily stable strategy* (ESS) which is "evolutionary" stable: once it is adopted by a population in a given environment, it cannot be invaded by any alternative strategy, initially rare. Many refinements to the ESS have been proposed, among them, one can cite a recent approach based on beliefs [8]. Evolution and promotion of cooperation have been extensively studied in the context of EGT. Thus, Wang et al. studied optimal interdependence between network nodes for the evolution of cooperation [9,10].

* Corresponding author. Tel.: +1 4186562131x3226; fax: +1 4186562324.
E-mail address: chaib@ift.ulaval.ca (B. Chaib-draa).

http://dx.doi.org/10.1016/j.amc.2015.07.068

		Player 2	
		C	D
Player 1	C	$2, \ 2$	$-1, \ 3$
	D	$3, -1$	$0, \ 0$

Fig. 1. The payoff matrix of prisoner's dilemma (PD).

Other researchers have studied the promotion of cooperation either by generating random variables that determine the social diversity of players engaging in the prisoner's dilemma game [4,11], by introducing coevolutionary rules [12], or by inducing appropriate payoff aspirations in a small-world networked game [13].

When dynamic games are being played by rational players, that is, selfish payoff maximizers, (and this paper is devoted to this sort of games) the collection of equilibrium strategies should possess the property of subgame perfection. In *subgame perfect Nash equilibrium* (SPNE), each agent's strategy has to be preferred by that player over any other strategy *in every possible situation* and not only in situations that can arise when every other player follows its respective equilibrium strategy. In other words, a SPNE is an equilibrium such that players' strategies constitute a Nash equilibrium in every subgame of the original game. Thus, a SPNE can be elaborated by *backward induction*, starting from the leaves players' actions and pulling up until the original game. Such backward induction eliminates *noncredible threats* which are tolerated by the Nash equilibria.

Computing SPNE is a difficult problem even for very simple forms of games, such as stage-games [14]. Often, these games and more generally dynamic games possess an infinity of equilibrium solutions and the set of solutions is usually much wider and very difficult to identify. Several successful attempts have been made to algorithmically identify a subset of, or all solutions in dynamic games. However, these attempts still have a number of important limitations. For example, the algorithm of [15] aims at finding a subset of pure action subgame-perfect equilibria (which do not always exist) in a repeated game. The algorithm of [16] aims at efficiently solving repeated games with no discounting but the assumption of no discounting makes trivial the task of identifying the set of solutions. On the other hand, the approach of [17,18] aims at computing correlated equilibria (not necessarily subgame-perfect) in stochastic games, which imposes additional restricting assumptions on the model, such as a presence of a constant and unlimited communication between players, or a mediating third party.

In this paper, we approach the problem of computing subgame-perfect equilibria in dynamic games with a complete information by first proposing a method to compute a nonempty subset of approximate subgame-perfect equilibria in any repeated game. We then demonstrate how to extend this method for approximating *all* equilibria in a repeated game, and solving more complex games, such as Markov chain games and stochastic games.

2. Example and motivation

A discussion of an example allows us to show the significance of repeated games and brings up the questions that we can address relative to these games. Probably the most famous example of a repeated game is prisoner's dilemma (PD), whose payoff matrix is shown in Fig. 1.

In this game, there are two players, called Player 1 and Player 2. At each stage of the repeated game, each player has a choice between two actions: *C* (for cooperation) and *D* (for defection, i.e., non-cooperation). When the two players simultaneously perform their actions, the resulting pair of actions induces a numerical payoff given by the payoff matrix. For example, if Player 1 plays action *C* and Player 2 plays action *D*, then the payoffs they obtain are respectively -1 and 3. The game then passes to the next stage, where it can be played again by the same pair of players.

As we can see, in one-shot interaction, the only outcome consistent with game theory prediction is (D, D) since each player is better off playing *D* whatever the other player does. On the other hand, if the game is repeated and the players value sufficiently future payoffs relative to the present ones, and if past actions are known, then (C, C) is an acceptable outcome for which no one wants to deviate. The reason is that if each player plays *C* as long as the other one has played *C* in the past, there is *an incentive* for both players to always play *C*. Indeed, the short-term outcome that can be obtained by playing *D* is more than offset by the future losses induced by always playing *D* at all future stages, that the opponent player can adopt as strategy of reprisal [19].

The utility function of the agent playing a repeated game is usually a non-decreasing function of accumulated payoffs. Game theory assumes that the goal of each player is to play rationally, i.e., to maximize its utility function. When the *a priori* information about all players' strategies and their real strategic preferences coincide, we refer to *equilibria*. A pair of "Tit-For-Tat" (TFT) strategies is a well-known example of equilibrium in the repeated prisoner's dilemma. TFT for Player *i*, where $i \in \{1, 2\}$, consists of starting the repeated PD game by playing action *C*. Then, Player *i* is supposed to play the same action as the very recent action played by its opponent. As explained earlier, a pair of TFT strategies constitutes an equilibrium in the repeated prisoner's dilemma if players are sufficiently patient, that is, they repeat the game with an utility function defined as a discounted sum of accumulated payoffs and the discount factor–γ–is close to 1.

Strategies such as TFT are called *non-stationary*, because they depend on the history of the repeated game. Therefore, an equilibrium strategy profile given by a pair of TFT strategies is a *non-stationary equilibrium* strategy profile. A stationary strategy, in turn, is a strategy that does not depend on the history of the repeated game. It is easy to verify that in the repeated prisoner's dilemma, a stationary equilibrium strategy profile is a pair of strategies that prescribe, to each player, to play action *D* at every stage of the repeated game. Indeed, if one player is supposed to always play *D*, it is rational, for the other player, to always play