Solving two-person zero-sum repeated games of incomplete information

Andrew Gilpin
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
gilpin@cs.cmu.edu

Tuomas Sandholm
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
sandholm@cs.cmu.edu

ABSTRACT

In repeated games with incomplete information, rational agents must carefully weigh the tradeoffs of advantageously exploiting their information to achieve a short-term gain versus carefully concealing their information so as not to give up a long-term informed advantage. The theory of infinitely-repeated two-player zero-sum games with incomplete information has been carefully studied, beginning with the seminal work of Aumann and Maschler. While this theoretical work has produced a *characterization* of optimal strategies, algorithms for *solving* for optimal strategies have not yet been studied. For the case where one player is *informed* about the true state of the world and the other player is *uninformed*, we provide a non-convex mathematical programming formulation for computing the value of the game, as well as optimal strategies for the informed player. We then describe an efficient algorithm for solving this difficult optimization problem to within arbitrary accuracy. We also discuss how to efficiently compute optimal strategies for the uninformed player using the output of our algorithm.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Miscellaneous; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: [Numerical Algorithms and Problems]

General Terms

Algorithms, Economics, Theory

Keywords

Computational game theory, equilibrium finding, two-person zero-sum incomplete-information repeated games, non-convex optimization

1. INTRODUCTION

An important topic in computational game theory is the study of algorithms for computing equilibrium strategies for games. Without such algorithms, the elegant game theory solution concepts would have little to offer in the way of guidance to designers and implementers of game-theoretic agents. On the contrary, equipping agents with these algorithms would enable them to use strategies that are determined by a game-theoretic analysis. In many games—including all finite two-person zero-sum games—such strategies are optimal for the agent regardless of the opponents' actions.

Most work on algorithms for equilibrium-finding has focused on the non-repeated setting in which a game is played only once. However, most agent interactions happen many times. For example, participants in a market will likely encounter the same buyers and sellers repeatedly (hence the importance of reputation). In this paper, we explicitly study a setting that models repeated interactions.

In addition to encountering one another multiple times, agents are typically endowed with some private information about the state of the world, that is, which game is being played. As does most prior work, we model this private information by treating the game as one of incomplete information. However, this is the first work to address computational considerations in repeated games of incomplete information, in which the agents' private information may be revealed over time—possibly inadvertently—by their actions.

One stream of related research on algorithms for repeated games falls under the category of *multiagent learning* (*e.g.*, [8, 13, 3, 11]) which is usually applied either when the rules of the game are unknown (*e.g.*, when a player is initially completely uninformed about the payoffs of the game) or when directly solving for good strategies is too difficult in a computational sense [20]. In the setting we study in this paper, both players know the rules of the game, and we demonstrate that it is not computationally infeasible to compute optimal strategies.

A closely related piece of work is due to Littman and Stone [14] who designed an algorithm for finding Nash equilibria in repeated games. That work studied the setting where both players know which stage game is being repeated. In our setting, the key difference is that only one of the players knows that.

In this paper we study a model of repeated interaction known as *two-person zero-sum repeated games of incomplete information*. We first review the necessary theory (Section 2) and present some illustrative examples (Section 3) of this class of games. Following that, for the case where one player is *informed* about the state of the world and the other player is *uninformed*, we derive a non-convex mathematical programming formulation for computing the value of the game (Section 4.1). This is a complicated optimization

Cite as: Solving two-person zero-sum repeated games of incomplete information, Andrew Gilpin and Tuomas Sandholm, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. XXX-XXX.

problems for which standard optimization algorithms do not apply. We describe and analyze a novel efficient algorithm for solving this problem to within arbitrary accuracy in Section 4.2. In Section 5.1 we demonstrate how the solution to our optimization problem yields an optimal strategy for the informed player. We also give an algorithm for the uninformed player to play optimally, in Section 5.2. Finally, in Section 6 we conclude and present some open problems.

2. PRELIMINARIES

In this section we review the definitions and concepts on which we will build. We first review in Section 2.1 some basic game theory for two-person zero-sum games with complete information (in which both players are fully informed about which game is being played). In Section 2.2 we review single-shot two-person zero-sum games with incomplete information (in which the players are only partially informed about the game being played). We conclude this section with the relevant theory of infinitely-repeated two-person zero-sum games with incomplete information (Section 2.3).

The material in this section is largely based on Aumann and Maschler's early work on the subject [1]. Myerson provides a textbook introduction [16] and Sorin provides a thorough treatment of two-person zero-sum repeated games [22].

2.1 Complete information zero-sum games

A two-person zero-sum game with complete information is given by $A \in \mathbb{Q}^{m\times n}$ with entries A_{ij}. In this game, player 1 plays an action in $\{1,\dots,m\}$ and player 2 *simultaneously* plays an action in $\{1,\dots,n\}$. If player 1 chooses i and player 2 chooses j, player 2 pays player 1 the amount A_{ij}. In general, the players may employ *mixed strategies* which are probability distributions over the available actions, denoted Δ_m and Δ_n, respectively, where

$$\Delta_m = \left\{ \mathbf{p} \in \mathbb{R}^m : \sum_{i=1}^{m} p_i = 1, \mathbf{p} \geq 0 \right\}$$

and similarly for Δ_n. If $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$, player 2 pays player 1 the quantity $\mathbf{x}A\mathbf{y}$ in expectation, where we take $\mathbf{x}$ to be a row vector and $\mathbf{y}$ to be a column vector.

Player 1, wishing to maximize the quantity $\mathbf{x}A\mathbf{y}$, while knowing that player 2 is a minimizer, faces the following optimization problem:

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}A\mathbf{y}. \tag{1}$$

Similarly, player 2 faces the problem

$$\min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}A\mathbf{y}. \tag{2}$$

The celebrated *minimax theorem* states that the values of these two problems are equal and can be simultaneously solved [23]. Hence, we may consider the problem of solving the following equation:

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}A\mathbf{y} = \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}A\mathbf{y}.$$

If $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ are solutions to the above problem then we say that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ are *minimax solutions* and we define the value of the game $v(A) = \bar{\mathbf{x}}A\bar{\mathbf{y}}$. (It is easy to see that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ also satisfy the weaker solution concept of Nash equilibrium, although in this paper we focus on the stronger minimax solution concept.)

We can reformulate player 1's problem (1) as the following linear program:

$$\max\{z : z\mathbf{1} - \mathbf{x}A \leq 0, \mathbf{x}\mathbf{1} = 1, \mathbf{x} \geq 0\} \tag{3}$$

where $\mathbf{1}$ is the all-ones column vector of appropriate dimension. Similarly, player 2's problem can be formulated as:

$$\min\{w : w\mathbf{1} - A\mathbf{y} \geq 0, \mathbf{1}^T\mathbf{y} = 1, \mathbf{y} \geq 0\}. \tag{4}$$

Thus, we can find minimax solutions for both players in polynomial time using linear programming.[1]

Later in this paper we also consider the *average* game which is given by a set of K possible games $\hat{A} = \{A^1, \dots, A^K\}$ with $A^i \in \mathbb{Q}^{m\times n}$, and a probability distribution $\mathbf{p} \in \Delta_K$. The possible games $\hat{A}$ are common knowledge, but neither player knows which game is actually being played. The actual game is chosen according to $\mathbf{p}$. Without knowing the actual game, but knowing the distribution, both players play strategies $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$. When the game A^i is drawn, player 2 pays player 1 $\mathbf{x}A^i\mathbf{y}$. Hence, the expected payment made in this game is

$$\sum_{i=1}^{K} p_i \mathbf{x}A^i\mathbf{y} = \sum_{i=1}^{K} \mathbf{x}(p_iA^i)\mathbf{y} = \mathbf{x}\left(\sum_{i=1}^{K} p_iA^i\right)\mathbf{y}.$$

Thus the average game is equivalent to playing the matrix game given by $A = \sum_{i=1}^{K} p_iA^i$ and we can compute the value of the average game using linear programming as applied to the matrix game A. We define the value of the matrix game as

$$v(\mathbf{p}, \hat{A}) = v\left(\sum_{i=1}^{K} p_iA^i\right).$$

2.2 Incomplete information zero-sum games

A two-person zero-sum game with incomplete information is given by matrices $A^{kl} \in \mathbb{Q}^{m\times n}$ for each $k \in \{1,\dots,K\}$ and $l \in \{1,\dots,L\}$. In this game, k is drawn according to some common-knowledge distribution $\mathbf{p} \in \Delta_K$ and the value k is communicated to player 1 (but not player 2). Similarly, l is drawn according to some common-knowledge distribution $\mathbf{q} \in \Delta_L$ and is communicated to player 2 only. Having learned their respective private values of k and l, both players play mixed strategies $\mathbf{x}^k$ and $\mathbf{y}^l$ yielding the expected payment $\mathbf{x}^kA^{kl}\mathbf{y}^l$ from player 2 to player 1. Given the probability distributions $\mathbf{p}$ and $\mathbf{q}$ and strategies $\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^K\}$ and $\mathbf{Y} = \{\mathbf{y}^1, \dots, \mathbf{y}^L\}$, the expected payment made in the game is

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{L} p_kq_lA^{kl}_{ij}x^k_iy^l_j.$$

There is a linear programming formulation for finding minimax solutions for two-person zero-sum games with incomplete information [18], but we do not describe it here as we do need it in this paper.

[1] We note that (3) and (4) are duals of each other and hence the strong duality theorem of linear programming can be used to prove the minimax theorem. See [5, pp. 230–233] for more details on the relationship between linear programming and zero-sum games.

2.3 Repeated incomplete information zero-sum games

A repeated two-person zero-sum game with incomplete information is defined by the same data as the single-shot setting described above, but the game is played differently. As before, k is drawn according to $\mathbf{p} \in \Delta_K$ and communicated to player 1 only, and l is drawn according to $\mathbf{q} \in \Delta_L$ and communicated to player 2 only. Now, however, the game A^{kl} is repeated. After each stage i, the players only observe each other's actions, s_1^i and s_2^i. In particular, the payment $A^{kl}_{s_1^i s_2^i}$ in each round i is not observed. (If it were, the players would quickly be able to "reverse engineer" the actual state of the world.)

The class of games described in the previous paragraph are two-person zero-sum repeated games *with lack of information on both sides*, since both players 1 and 2 are uninformed about some aspect of the true state of the world. In this paper, we limit ourselves to two-person zero-sum repeated games *with lack of information on one side*. In this setting, player 1 (the *informed player*) is told the true state of the world. Player 2, the *uninformed player*, is not told anything, *i.e.,* we assume $|L| = 1$. We must consider this more specific class of games for several reasons.

- The first reason is that the notion of *value* in games with lack of information on both sides is unclear. There are two approaches to studying the repeated games in this model. The first is to consider the n-stage game, denoted Γ_n, and its value $v(\Gamma_n)$ (which clearly exists since Γ_n is finite), and then determine $\lim_{n\to\infty} v(\Gamma_n)$, if it exists. (Observe that Γ_1 is the same as the game described in Section 2.2.) The second approach is to consider the infinitely-repeated game, denoted Γ_∞, and determine its value $v(\Gamma_\infty)$ directly, if it exists. In two-person zero-sum repeated games with lack of information on one side, both $\lim_{n\to\infty} v(\Gamma_n)$ and $v(\Gamma_\infty)$ exist and are equal, so either approach is suitable [1]. However, there are games with lack of information on both sides for which $v(\Gamma_\infty)$ does not exist (although it is known that $\lim_{n\to\infty} v(\Gamma_n)$ exists [15]).

- The second reason involves the choice of modeling *payoffs* in repeated games with incomplete information. As discussed above, the players do not observe the payoffs in each stage. How then does one even evaluate a player's strategy or consider the notion of value? It is customary to consider the mean-payoff, so we say that after the N-th stage, player 1 receives a payoff

  $$\frac{1}{N}\sum_{i=1}^{N} A^{kl}_{s_1^i s_2^i}$$

  from player 2.

  As an alternative to the mean-payoff model, one could also study the seemingly more natural λ-discounted game Γ_λ as $\lambda \to 0$. This game is repeated infinitely many times and player 1 receives the payoff

  $$\sum_{i=1}^{\infty} \lambda(1-\lambda)^{i-1} A^{kl}_{s_1^i s_2^i}$$

  from player 2. It can be shown that

  $$v(\Gamma_\infty) = \lim_{n\to\infty} v(\Gamma_n) = \lim_{\lambda\to 0} v(\Gamma_\lambda)$$

  where $v(\Gamma_\infty)$ exists [22, Lemma 3.1]. Since, as discussed in the previous point, this condition holds for games with lack of information on one side, we have that the value of the game is the same regardless of which payoff metric we choose.

  Given this flexibility in choice of payoff measure, we simply restrict ourselves to studying the mean-payoff game for computational reasons.

- Finally, as we will elaborate in Section 4, our optimization approach for computing optimal strategies in the game depends heavily on a characterization of equilibrium strategies that unfortunately only holds for games with lack of information on one side.

In summary, we limit ourselves to the one-sided lack of information setting for various conceptual and computational reasons. Developing concepts and algorithms for the more general two-sided setting is an important area of future research which we discuss further in Section 6.

3. EXAMPLES

Before describing our algorithms for computing both the value and optimal strategies for games in our model, we discuss some classical examples that illustrate the richness and complexity of our model.

In any repeated game with incomplete information, the players must take into account to what extent their actions reveal their private information, and to what extent this revelation will affect their future payoffs. In the following three subsections, we present three examples where the amount of information revelation dictated by the optimal strategy differs. The first two examples are due to Aumann and Maschler [1] and the third is due to Zamir [25].

3.1 Completely unrevealing strategy

Our first example is given in Figure 1. There are two possible states of the world, A and B, and each is chosen by nature with probability 0.5. The true state of the world is communicated to player 1.

State A

	L	R
U	1	0
D	0	0

State B

	L	R
U	0	0
D	0	1

Figure 1: The stage games for the unrevealing example. If the state of the world is State A, then the game on the left is played. Otherwise, the game on the right is played.

Consider what happens if player 1, after being informed that the true state is A, plays U every time. (This is a weakly dominant strategy for player 1 in the state A game.) Eventually it will occur to player 2, who is observing player 1's actions, that player 1 is playing U because the players are in state A and player 1 is hoping to get the payoff of 1. Observing this, player 2 will switch to playing R, guaranteeing a payoff of 0.

A similar line of reasoning in the case of state B appears to demonstrate that player 1 can only achieve a long-term payoff of 0 as player 2 will eventually figure out what actual game is being played. However, somewhat unintuitively, consider what happens if player 1 *ignores* her private signal. Then no matter what strategy player 1 uses, player 2 will not be able to infer the game being played. In fact, it as if the players are playing the *average* game in Figure 2. In this game, both players' (unique) optimal strategy is to play both of their actions with probability 0.5, for an expected payoff of 0.25. Thus player 1 achieves an average expected payoff of 0.25, compared to the payoff of 0 that she would get if player 2 were able to infer the actual game from player 1's action history. Note that player 1's strategy is *completely non-revealing* since player 2 will never be able to determine the true state of the world based on player 1's actions.

	L	R
U	1/2	0
D	0	1/2

Figure 2: The *average* game corresponding to the case where player 1 ignores her private information in the game in Figure 1.

Although we do not prove it here, the above strategy for player 1 is optimal. Intuitively, if player 1 were to slightly alter her strategy to take advantage of her private information, player 2 would observe this and would then deviate to the unique best response to player 1's altered strategy. Thus, any advantage player 1 could possibly get would be short-term, and would not be nearly enough to compensate for the long-term losses that player 2 would be able to inflict.

3.2 Completely revealing strategy

Our second example is given in Figure 3. Again, there are two possible states of the world, A and B, and each is chosen by nature with probability 0.5. The true state of the world is communicated to player 1.

State A

	L	R
U	-1	0
D	0	0

State B

	L	R
U	0	0
D	0	-1

Figure 3: The stage games for the revealing example. If the state of the world is State A, then the game on the left is played. Otherwise, the game on the right is played.

Here, a payoff of 0 is clearly the best player 1 can hope for, and this outcome can be achieved using the following strategy: "Always play D if the state of the world is A; otherwise always play U". No matter what strategy player 2 uses, player 1 obtains a payoff of 0. Since this is the highest possible payoff, this is clearly an optimal strategy. Note that this strategy is *completely revealing* since player 2 will be able to determine the true state of the world.

3.3 Partially revealing strategy

Our third example is given in Figure 4. Again, there are two possible states of the world, A and B, and each is chosen by nature with probability 0.5. The true state of the world is communicated to player 1.

If, as in the first example, player 1 completely ignores her private information, the game reduces to the one in Figure 5. In this case, player 2 can always play R and thus guarantee a payout of at most 0.

State A

	L	M	R
U	4	0	2
D	4	0	-2

State B

	L	M	R
U	0	4	-2
D	0	4	2

Figure 4: The stage games for the partially-revealing example. If the state of the world is State A, then the game on the left is played. Otherwise, the game on the right is played.

State A

	L	M	R
U	2	2	0
D	2	2	0

Figure 5: The *average* game corresponding to the case where player 1 ignores her private information in the game in Figure 4.

If, on the other hand, player 1 completely reveals her private information, then player 2 will have the following optimal strategy: "Always play M if the inferred state is A; otherwise always play L". Again, player 2 is able to guarantee a maximum payout of 0.

Suppose player 1 now employs the following strategy: "If the state of the world is A, then with probability 0.75 always play U, otherwise always play D; if the state of the world is B, then with probability 0.25 always play U, otherwise always play D". Suppose further that player 2 knows this strategy. If player 2 observes that player 1 is always playing U, then player 2 can infer that

$$\begin{aligned} Pr[A|U] &= \frac{Pr[U|A]Pr[A]}{Pr[U]} \\ &= \frac{Pr[U|A]Pr[A]}{Pr[U,A]+Pr[U,B]} \\ &= \frac{0.75 \cdot 0.5}{0.5 \cdot 0.75 + 0.5 \cdot 0.25} \\ &= \frac{3}{4}. \end{aligned}$$

Hence, player 2 is faced with the decision problem

	L	M	R
U	3	1	1

and so can do by best by achieving a payout of 1. A similar computation shows the same is true if player 2 observes player 1 to always be playing D. Therefore, player 1 achieves a payoff of 1, which is better than she would have done had she either completely revealed her private information or completely ignored her private information. By *partially revealing* this information, she has boosted her payoff.

4. OPTIMIZATION FORMULATION

In this section, we review Aumann and Maschler's theorem for characterizing the value of two-person zero-sum repeated games with lack of information on one side. Unfortunately, this theorem does not include an algorithm for actually computing the value. Using this characterization we derive a non-convex optimization problem for computing the value of the game (Section 4.1). Non-convex optimization problems are in general $\mathcal{NP}$-complete [9] so there is little hope of employing a general-purpose algorithm for solving non-convex optimization problems. Instead, we give a specialized algorithm that computes the value of the game to within additive error ϵ for any given target accuracy $\epsilon > 0$ (Section 4.2). For games with a constant number of world states (but with a non-constant number of actions available to the players) we can compute such a solution in time polynomial in the number of actions and $\frac{1}{\epsilon}$.

4.1 Optimization formulation derivation

Consider the two-person zero-sum game with incomplete information given by matrices $A^k \in \mathbb{Q}^{m \times n}$ for $k \in \{1, \ldots, K\}$, and let $\mathbf{p}^* \in \Delta_K$ be the probability with which k is chosen and communicated to player 1. (This is a game with lack of information on one side only, so player 2 does not receive any information about the true state of the world.) Denote the infinitely-repeated version of this game as Γ_∞. We are interested in computing $v(\Gamma_\infty)$.

Recall the *average game* which for some $\mathbf{p} \in \Delta_K$ has payoff matrix

$$A(\mathbf{p}) = \sum_{i=1}^{K} p_i A^i.$$

As discussed in Section 2.1, the value of this game is $v(\mathbf{p}, A) = v(A(\mathbf{p}))$ and can be computed using LP. In what follows we omit A when it can be inferred from context, and instead simply discuss the value function $v(\mathbf{p})$.

Consider now the *concavification* of $v(\mathbf{p})$ with respect to $\mathbf{p}$, that is, the point-wise smallest (with respect to $v(\cdot)$) concave function that is greater than v for all $\mathbf{p} \in \Delta_K$. Letting $v' \geq v$ denote that $v'(\mathbf{p}) \geq v(\mathbf{p})$ for all $\mathbf{p} \in \Delta_K$, we can formally write

$$\mathbf{cav}\ v(\mathbf{p}) = \inf_{v'} \left\{ v'(\mathbf{p}) : v' \text{ concave}, v' \geq v \right\}.$$

Aumann and Maschler's surprising and elegant result [1] states that $v(\Gamma_\infty) = \mathbf{cav}\ v(\mathbf{p})$. Our goal of computing $v(\Gamma_\infty)$ can thus be achieved by computing $\mathbf{cav}\ v(\mathbf{p})$.

A basic result from convex analysis [4] shows that the convex hull of an n-dimensional set S can be formed by taking convex combinations of $n+1$ points from S. Hence, the K-dimensional point $(\mathbf{p}, \mathbf{cav}\ v(\mathbf{p}))$ can be represented as the convex combination of $K+1$ points $(\mathbf{p}^i, v(\mathbf{p}^i))$, $i \in \{1, \ldots, K+1\}$. (Note that $\mathbf{p}$ is $(K-1)$-dimensional, not K-dimensional.) In particular, for any $(\mathbf{p}, \mathbf{cav}\ v(\mathbf{p}))$, there exists $\alpha \in \Delta_{K+1}$ and points

$$\left\{ \left(\mathbf{p}^1, v\left(\mathbf{p}^1\right)\right), \ldots, \left(\mathbf{p}^{K+1}, v\left(\mathbf{p}^{K+1}\right)\right) \right\}$$

such that

$$\mathbf{p} = \sum_{i=1}^{K+1} \alpha_i \mathbf{p}^i$$

and

$$\mathbf{cav}\ v(\mathbf{p}) = \sum_{i=1}^{K+1} \alpha_i v(\mathbf{p}^i).$$

Hence, we can rewrite the problem of computing $\mathbf{cav}\ v(\mathbf{p})$ as the following optimization problem:

$$\begin{array}{lll} & \max & \displaystyle\sum_{i=1}^{K+1} \alpha_i v(\mathbf{p}^i) \\ \text{(P1)} & \text{such that} & \displaystyle\sum_{i=1}^{K+1} \alpha_i \mathbf{p}^i = \mathbf{p} \\ & & \mathbf{p}^i \in \Delta_K \text{ for } i \in \{1, \ldots, K+1\} \\ & & \alpha \in \Delta_{K+1} \end{array}$$

A solution to Problem (P1) therefore yields the value of Γ_∞. Unfortunately, this does not immediately suggest a good algorithm for solving this optimization problem. First, the optimization problem depends on the quantities $v(\mathbf{p}^i)$ for variables $\mathbf{p}^i$. As discussed in Section 2.1, the value $v(\mathbf{p}^i)$ is itself the solution to an optimization problem (namely a linear program), and hence a closed-form expression is not readily available. Second, the first constraint is non-linear and non-convex. Continuous optimization technology is much better suited for convex problems [17], and in fact non-convex problems are $\mathcal{NP}$-complete in general [9]. In the following subsection, we present a numerical algorithm for solving Problem (P1) with arbitrary accuracy ϵ.[2]

4.2 Solving the formulation

Our algorithm is closely related to *uniform grid methods* which are often used for solving extremely difficult optimization problems when no other direct algorithms are available [17]. Roughly speaking, these methods discretize the feasible space of the problem and evaluate the objective function at each point. Our problem differs from the problems normally solved via this approach in two ways. The first is that the feasible space of Problem (P1) has additional structure not normally encountered. Most uniform grid methods have a hyper-rectangle as the feasible space. In contrast, our feasible space is the product of several simplices of different dimension, which are related to each other via a non-convex equality constraint (the first constraint in Problem (P1)). Second, evaluating the objective function of Problem (P1) is not straightforward as it depends on several values of $v(\mathbf{p}^i)$ which are themselves the result of an optimization problem.

As above we consider a game given by K matrices A^k, but now for simplicity we require the rational entries A^k_{ij} to be in the unit interval $[0, 1]$. This is without loss of generality since player utilities are invariant with respect to positive affine transformations and so the necessary rescaling does not affect their strategies.

[2] Unfortunately, our algorithm cannot solve Problem (P1) exactly, but rather only to within additive error ϵ. However, this appears unavoidable since there are games for which the (unique) optimal play of player 1 involves probabilities that are irrational numbers. An example of such a game is due to Aumann and Maschler [1, p. 79–81]. This immediately implies that there does not exist a linear program whose coefficients are arithmetically computed from the problem data whose solution yields an optimal strategy to Problem (P1) since every linear program with rational coefficients has a rational solution. Furthermore, any numerical algorithm will not be able to compute an exact solution for similar reasons. Note that this is very different from the case of two-person zero-sum games in which optimal strategies consisting of rational numbers as probabilities always exist.

Our algorithm for solving Problem (P1) within additive error ϵ proceeds as follows.

Procedure SolveP1

1. Let $C = \lceil \frac{1}{\epsilon} \rceil$.
2. Let $\mathbf{Z}(C) = \left\{ \left(\frac{d_1}{C}, \ldots, \frac{d_K}{C} \right) : d_i \in \mathbb{N}, \sum_{i=1}^{K} d_i = C \right\}$. Denote the points of $\mathbf{Z}(C)$ as $\mathbf{p}^1, \ldots, \mathbf{p}^{|\mathbf{Z}(C)|}$.
3. For each point $\mathbf{p}^i \in \mathbf{Z}(C)$, compute and store $v(\mathbf{p}^i)$ using linear program (3).
4. Solve linear program (P2):

$$
\text{(P2)} \qquad \begin{array}{ll}
\max & \displaystyle\sum_{i=1}^{|\mathbf{Z}(C)|} \alpha_i v(\mathbf{p}^i) \\
\text{such that} & \displaystyle\sum_{i=1}^{|\mathbf{Z}(C)|} \alpha_i \mathbf{p}^i \leq \mathbf{p} \\
 & \alpha \in \Delta_{|\mathbf{Z}(C)|}
\end{array}
$$

5. Output the value of linear program (P2) as an approximation of the value of the game.

We now analyze the above algorithm. We first recall the following classic fact about the value function $v(\mathbf{P})$.

Lemma 1. *$v(\mathbf{p})$ is Lipschitz with constant 1, that is,*

$$|v(\mathbf{p}) - v(\mathbf{p}')| \leq \|\mathbf{p} - \mathbf{p}'\|_\infty.$$

Proof. This is immediate in our case since we assume that all entries of A are in the range $[0,1]$. □

Before presenting our main theorem, we present a technical lemma which shows that our discretization (*i.e.,* our design of $\mathbf{Z}(C)$, is sufficiently refined. The proof is immediate from the definition of $\mathbf{Z}(C)$ and is omitted.

Lemma 2. *Let $\mathbf{p} \in \Delta_K$, $\epsilon > 0$, and $C = \lceil \frac{K}{\epsilon} \rceil$. There exists $\mathbf{q} \in \mathbf{Z}(C)$ such that*

$$\|\mathbf{p} - \mathbf{q}\|_\infty \leq \frac{1}{C}.$$

Theorem 1. *Let $v(\Gamma_\infty)$ be the value of the infinitely-repeated game and let v^* be the value output by the above algorithm with input $\epsilon > 0$. Then*

$$v(\Gamma_\infty) - v^* \leq \epsilon.$$

Proof. We only need to show that there exists a feasible solution to the linear program (P2) whose objective value satisfies the inequality (the optimal answer to the linear program could be even better). Let $\bar{\alpha}, \bar{\mathbf{p}}^1, \ldots, \bar{\mathbf{p}}^{K+1}$ be optimal solutions to (P1). We construct a feasible solution α to linear program (P2) as follows. For each $i \in \{1, \ldots, K+1\}$, choose $\mathbf{p}^j \in \mathbf{Z}(C)$ such that $\|\bar{\mathbf{p}}^i - \mathbf{p}^j\|_\infty$ is minimized (breaking ties arbitrarily) and set $\alpha_j = \bar{\alpha}_i$. Assign $\eta(i) = j$. Leave all other entries of α zero. Let $N = \{i : \alpha_i > 0\}$ be the index set for the positive entries of α and let

$$v = \sum_{i=1}^{|\mathbf{Z}(C)|} \alpha_i v(\mathbf{p}^i).$$

Clearly, this is a lower bound on the objective value of linear program (P2). Now we can write:

$$
\begin{aligned}
v(\Gamma_\infty) - v &= \sum_{i=1}^{K+1} \bar{\alpha}_i v\left(\bar{\mathbf{p}}^i\right) - \sum_{i=1}^{|\mathbf{Z}(C)|} \alpha_i v\left(\mathbf{p}^i\right) \\
&= \sum_{i=1}^{K+1} \bar{\alpha}_i v\left(\bar{\mathbf{p}}^i\right) - \sum_{i \in N} \alpha_i v\left(\mathbf{p}^i\right) \\
&= \sum_{i=1}^{K+1} \bar{\alpha}_i v\left(\bar{\mathbf{p}}^i\right) - \sum_{i=1}^{K+1} \alpha_{\eta(i)} v\left(\mathbf{p}^{\eta(i)}\right) \\
&= \sum_{i=1}^{K+1} \bar{\alpha}_i v\left(\bar{\mathbf{p}}^i\right) - \sum_{i=1}^{K+1} \bar{\alpha}_{\eta(i)} v\left(\mathbf{p}^{\eta(i)}\right) \\
&= \sum_{i=1}^{K+1} \bar{\alpha}_i \left[v\left(\bar{\mathbf{p}}^i\right) - v\left(\mathbf{p}^{\eta(i)}\right) \right] \\
&\leq \sum_{i=1}^{K+1} \bar{\alpha}_i \left\| \bar{\mathbf{p}}^i - \mathbf{p}^{\eta(i)} \right\|_\infty \\
&\leq \sum_{i=1}^{K+1} \bar{\alpha}_i \frac{1}{C} = \frac{1}{C} \leq \epsilon
\end{aligned}
$$

The first inequality is by Lemma 1, the second inequality is by Lemma 2, and the third inequality is by the definition of C. □

We now analyze the time complexity of our algorithm. We first prove a simple lemma.

Lemma 3. *For $C = \lceil \frac{1}{\epsilon} \rceil$ the set*

$$\mathbf{Z}(C) = \left\{ \left(\frac{d_1}{C}, \ldots, \frac{d_K}{C} \right) : d_i \in \mathbb{N}, \sum_{i=1}^{K} d_i = C \right\}$$

defined in step 2 of the above algorithm satisfies

$$|\mathbf{Z}(C)| = \binom{C + K - 1}{K - 1} \leq (C+1)^K.$$

Proof. The equality is a simple exercise in combinatorics and the inequality is standard. □

We analyze our algorithm in terms of the number of linear programs it solves. Note that each linear program is solvable in polynomial-time, *e.g.*, by the ellipsoid method [12] or by interior-point methods [24]. Step 3 of the algorithm clearly makes $|\mathbf{Z}(C)|$ calls to a linear program solver. By Lemma 3, we have $|\mathbf{Z}(C)| \leq (C+1)^K$. Each of these linear program has $m+1$ variables and $n+1$ constraints, where m and n are the numbers of actions each player has in each stage game.

Similarly, the linear program solved in step 4 of the algorithm also has at most $|\mathbf{Z}(C)| \leq (C+1)^K$ variables. Hence we have:

Theorem 2. *The above algorithm solves $(C+1)^K$ linear program that are of size polynomial in the size of the input data, and solves one linear program with $(C+1)^K$ variables.*

Therefore, for a fixed number of possible states K, our algorithm runs in time polynomial in the number of actions available to each player and in $\frac{1}{\epsilon}$.

5. FINDING THE PLAYERS' STRATEGIES

In this section, we demonstrate how we can use the output of our algorithm to construct the player's strategies. The strategy of player 1 (the informed player) can be constructed explicitly from the values of the variables in the linear program (P2) solved in our algorithm. Player 2 (the uninformed player) does not have such an explicitly represented strategy. However, we can use existing approaches (in conjunction with the output from our algorithm) to describe a simple algorithmic procedure for player 2's strategy.

5.1 The informed player's strategy

As alluded to in the examples, player 1's optimal strategy is of the following form. Based on the revealed choice of nature, player 1 performs a type-dependent lottery to select some distribution $\mathbf{q} \in \Delta_K$. Then she always plays as if the stage game were the average game induced by a distribution that does not reveal the true state of the world to player 2.

Let $\alpha, \mathbf{p}^1, \ldots, \mathbf{p}^{K+1}$ be solutions to Problem (P1). The following strategy is optimal for player 1 [1].

> Let $k \in \{1, \ldots, K\}$ be the state revealed (by nature) to player 1. Choose $i \in \{1, \ldots, K+1\}$ with probability
>
> $$\frac{\alpha_i p_k^i}{p_k}$$
>
> where $\mathbf{p} = \{p_1, \ldots, p_k\}$ is the probability that natures chooses state k. Play the mixed equilibrium strategy corresponding to the average game given by distribution $\mathbf{p}^i$ in every stage.

Thus the informed player is using her private information once and for all at the very beginning of the infinitely-repeated game, and then playing always as if the game were actually the average game induced by the distribution $\mathbf{p}^i$. The strength of this strategy lies in the fact that player 2, after observing player 1's strategy, is unable to determine the actual state of nature *even after learning which number i player 1 observed in her type-dependent lottery.*

This surprisingly and conceptually simple strategy immediately suggests a similar strategy for player 1 based on the output of our algorithm. Let α be a solution to linear program (P2) solved during the execution of our algorithm, let $\{\mathbf{p}^1, \ldots, \mathbf{p}^{|\mathbf{Z}(C)|}\} = \mathbf{Z}(C)$, and let k be the actual state of nature. The strategy is as follows: "Choose $i \in \{1, \ldots, |\mathbf{Z}(C)|\}$ with probability

$$\frac{\alpha_i p_k^i}{p_k}.$$

Play a mixed equilibrium strategy to the average game corresponding to the distribution $\mathbf{p}^i$ in every stage thereafter". Using reasoning completely analogous to the reasoning of Aumann and Maschler [1], this strategy guarantees player 1 a payoff of at least $v(\Gamma_\infty) - \epsilon$. (If we were able to solve problem (P1) optimally then we would have $\epsilon = 0$.)

5.2 The uninformed player's strategy

We now describe how player 2's strategy can be constructed from the solution to our algorithm. Unlike in the case of player 1, there is no concise, explicit representation of the strategy. Rather the prescription is in the form of an algorithm.

The driving force behind this technique is Blackwell's approachability theory [2] which applies to games with vector payoffs.[3] The basic idea is that player 2, instead of attempting to evaluate her expected payoff, instead considers her *vector payoff*, and then attempts to force this vector payoff to *approach* some set.

Because $\mathbf{cav}\ v(\mathbf{p})$ is concave, there exists $\mathbf{z} \in \mathbb{R}^K$ such that

$$\sum_{i=1}^{K} p_i z_i = \mathbf{cav}\ v(\mathbf{p})$$

and

$$\sum_{i=1}^{K} q_i z_i \geq \mathbf{cav}\ v(\mathbf{q}), \quad \forall \mathbf{q} \in \Delta_K.$$

Let $S = \left\{\mathbf{s} \in \mathbb{R}^K | \mathbf{s} \leq \mathbf{z}\right\}$. Blackwell's approachability theorem states that there exists a strategy for player 2 such that for *any* strategy of player 1, player 2 can receive a vector payoff arbitrarily close (in a precise sense) to the set S. Since S can be interpreted as the set of affine functions majorizing $v(\mathbf{p})$ (and, hence, majorizing $\mathbf{cav}\ v(\mathbf{p})$), then player 2 can force a payout arbitrarily close to the payoff that player 1 guarantees.

Following from the above discussion, we can state the following optimal strategy for player 2. For each stage n and each $i \in K$, let u_n^i be player 2's payout to player 1 (given that the state of the world is actually i). Now define

$$w_n^i = \frac{\sum_{j=1}^{n} u_j^i}{n}$$

to be player 2's average payoff vector to player 1. In stage 1 and in any stage n where $\mathbf{w}_n = (w_n^1, \ldots, w_n^K) \in S$, let player 2 play an arbitrary action (this is acceptable since so far player 2 is doing at least as well as possible). At stage n where $\mathbf{w}_n \notin S$, let player 2 choose her move according to a distribution $\mathbf{y}$ satisfying

$$\min_{\mathbf{y} \in \Delta_n} \max_{\mathbf{x} \in \Delta_m} \sum_{i=1}^{K} \left(w_{n-1}^i - \varepsilon^i(\mathbf{w}_{n-1}) \right) \mathbf{x} A^i \mathbf{y} \tag{5}$$

where $\varepsilon(\mathbf{w}_{n-1})$ is the (unique) point in S that is closest to $\mathbf{w}_{n-1}$. Blackwell's approachability theorem [2] states that player 2's vector payoff converges to S regardless of player 1's strategy.

The above discussion thus shows how player 2, using the information output by our algorithm, can be used to generate a strategy achieving the optimal payoff. In each stage, at most all that is required is solving an instance of Equation 5, which can be solved in polynomial time using linear programming.

6. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we studied computational approaches for finding optimal strategies in repeated games with incomplete information. In such games, an agent must carefully weigh the tradeoff between exploiting its information to achieve

[3] We do not include a full description of this theory here. Complete descriptions are provided by Myerson [16, pp. 357–360], Sorin [22, Appendix B], or Blackwell's original paper [2].

a short-term gain versus carefully concealing its information so as not to give up a long-term informed advantage. Although the theoretical aspects of these games have been studied, this is the first work to develop algorithms for *solving* for optimal strategies. For the case where one player is *informed* about the true state of the world and the other player is *uninformed*, we derived a non-convex mathematical programming formulation for computing the value of the game, as well as optimal strategies for the informed player. We then described an algorithm for solving this difficult optimization problem to within arbitrary accuracy. We also described a method for finding the optimal strategy for the uninformed player based on the output of the algorithm.

Directions for future research are plentiful. This paper has only analyzed the case of one-sided information. Developing algorithms for the case of lack of information on both sides would be an interesting topic. However, this appears difficult. For one, the notion of the value of the game is less well understood in these games. Furthermore, there is no obvious optimization formulation that models the equilibrium problem (analogous to Problem (P1)). Other possible directions include extending this to non-zero-sum games [10] as well as to games with many players. Again, these tasks appear difficult as the characterizations of equilibrium strategies become increasingly complex.

Yet another possible algorithmic approach would be to tackle the problem via Fenchel duality (Rockafellar [19] is a standard reference for this topic). Fenchel duality has been employed as an alternative method of proving various properties about repeated games [6, 7]. In particular, the Fenchel biconjugate of $v(\mathbf{p})$ yields $\mathbf{cav}\ v(\mathbf{p})$. Given the close relationship between Fenchel duality and optimization theory, an intriguing possibility would be to use Fenchel duality to derive an improved optimization algorithm for the problem studied in this paper.

The class of games we study is a special case of the more general class of *stochastic games* [21]. That class of games allows for a much richer signaling structure (rather than the limited signaling structure we consider in which only the players' actions are observable), as well as transitioning to different stage games based on the choices of the players and possible chance moves. Developing a solid algorithmic understanding of the issues in those richer games is another important area of future research.

7. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under ITR grant IIS-0427858.

8. REFERENCES

[1] R. J. Aumann and M. Maschler. *Repeated Games with Incomplete Information.* MIT Press, 1995. With the collaboration of R. Stearns. This book contains updated versions of four papers originally appearing in *Report of the U.S. Arms Control and Disarmament Agency*, 1966–68.

[2] D. Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[3] R. Brafman and M. Tennenholtz. A near-optimal polynomial time algorithm for learning in certain classes of stochastic games. *Artificial Intelligence*, 121:31–47, 2000.

[4] C. Carathéodory. Uber den Variabiletätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico de Palermo*, 32:193–217, 1911.

[5] V. Chvátal. *Linear Programming.* W. H. Freeman and Company, New York, NY, 1983.

[6] B. De Meyer. Repeated games, duality and the central limit theorem. *Mathematics of Operations Research*, 21:237–251, 1996.

[7] B. De Meyer and D. Rosenberg. "Cav u" and the dual game. *Mathematics of Operations Research*, 24:619–626, 1999.

[8] D. Fudenberg and D. Levine. *The Theory of Learning in Games.* MIT Press, 1998.

[9] M. Garey and D. Johnson. *Computers and Intractability.* W. H. Freeman and Company, 1979.

[10] S. Hart. Nonzero-sum two-person repeated games with incomplete information. *Mathematics of Operations Research*, 10:117–153, 1985.

[11] J. Hu and M. P. Wellman. Nash Q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4:1039–1069, 2003.

[12] L. Khachiyan. A polynomial algorithm in linear programming. *Soviet Math. Doklady*, 20:191–194, 1979.

[13] M. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Int. Conf. on Machine Learning (ICML)*, pages 157–163, 1994.

[14] M. Littman and P. Stone. A polynomial-time Nash equilibrium algorithm for repeated games. In *Proc. of the ACM Conference on Electronic Commerce (ACM-EC)*, pages 48–54, San Diego, CA, 2003.

[15] J.-F. Mertens and S. Zamir. The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, 1:39–64, 1971.

[16] R. Myerson. *Game Theory: Analysis of Conflict.* Harvard University Press, Cambridge, 1991.

[17] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers, 2004.

[18] J.-P. Ponssard and S. Sorin. The L-P formulation of finite zero-sum games with incomplete information. *International Journal of Game Theory*, 9:99–105.

[19] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[20] T. Sandholm. Perspectives on multiagent learning. *Artificial Intelligence*, 171:382–391, 2007.

[21] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39:1095–1100, 1953.

[22] S. Sorin. *A First Course on Zero-Sum Repeated Games.* Springer, 2002.

[23] J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

[24] S. J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, PA, 1997.

[25] S. Zamir. Repeated games of incomplete information: Zero-sum. In R. J. Aumann and S. Hart, editors, *Handbook of Game Theory, Vol. I*, pages 109–154. North Holland, 1992.