

A Framework for Sequential Planning in Multi-Agent Settings

Piotr J. Gmytrasiewicz and Prashant Doshi*

Department of Computer Science
University of Illinois at Chicago
{piotr, pdoshi}@cs.uic.edu

Abstract

This paper extends the framework of partially observable Markov decision processes (POMDPs) to multi-agent settings by incorporating the notion of agent models, or types defined in games of incomplete information, and by using Bayesian update over models during repeated interactions. We allow agents to have beliefs over physical states of the environment and over models of other agents which could include their belief states. Our approach complements a more traditional approach in stochastic games that uses equilibria as a solution paradigm. Our work seeks to avoid some of the drawbacks of game-theoretic equilibria which may be non-unique and unable to capture off-equilibrium behaviors. Our approach does so at the cost of having to represent, process and continually revise models of other agents. Agents' beliefs are, in general, arbitrarily nested, and optimal solutions to decision making problems are only asymptotically computable. However, approximate belief updates and approximately optimal plans are computable.

1 Introduction

We develop a framework for sequential rationality of autonomous agents interacting with other agents within a common, and possibly uncertain, environment. We use the normative paradigm of decision-theoretic planning under uncertainty as represented by partially observable Markov decision processes (POMDPs) [6, 16, 27]. We make this framework applicable to agents interacting with other agents by allowing them to have beliefs not only about the physical environment, but also about the other agents; i.e., their abilities, sensing capabilities, beliefs, preferences, and intentions.

The formalism of Markov decision processes has been extended to multiple agents before, giving rise to stochastic games or Markov games [10, 25]. Traditionally, the solution concept used for stochastic games is that of Nash equilibria. Some recent work in AI follows that tradition [5, 14, 17, 18]. However, while Nash equilibria are useful for describing a multi-agent system when, and if, it has reached a stable state, this solution concept is not sufficient as a general control paradigm. The main reasons are that there may be multiple equilibria with no clear way to choose among them (non-uniqueness), and the fact that equilibria do not specify actions in cases in which agents believe that other agents may act not according to their equilibrium strategies (incompleteness) [4, 15].

Other extensions of POMDPs to multiple agents appeared in [3, 30]. They have been called decentralized POMDPs (DEC-POMDPs), and are related to decentralized control problems [24]. DEC-POMDP framework assumes that the agents are fully cooperative, i.e., they have common reward function and form a team. Furthermore, it is assumed that the optimal joint policy is computed centrally and then distributed among the agents, which makes it a variant of multibody planning [27].

*This research is supported by the National Science Foundation CAREER award IRI-9702132, and NSF award IRI-0119270.

Our formalism, called interactive POMDPs (I-POMDPs) is applicable to autonomous agents with possibly conflicting objectives, operating in partially observable environments, who locally compute what actions they should execute to optimize their preferences given what they believe. We are motivated by a subjective approach to probability in games [15], and we combine POMDPs, Bayesian games, work on interactive belief systems [1, 12, 19], and related work [4, 8]. The unique aspect of I-POMDPs is that they prescribe action based on an agent’s beliefs about other agents and about their expected behaviors. This generalizes and complements the traditional equilibrium approach [15] – if the agent believes that others will act according to an equilibrium then it also chooses to act out its own part of this equilibrium. However, if the agent believes that others will diverge from an equilibrium behavior then the agent can still optimize. This approach, also called decision-theoretic approach to game theory [21], is capable of avoiding the difficulties of non-uniqueness and incompleteness of traditional equilibrium approach, but at the cost of processing and maintaining possibly infinitely nested interactive belief systems [1, 2] (also called knowledge structures [28].) As a result, only approximate belief updates and approximately optimal solutions to optimal planning problems are computable in general.

Our approach follows a tradition of knowledge-based paradigm of agent design, according to which it is useful for agents to represent and reason with the important elements of their environment to allow them to function efficiently. We extend this paradigm to other agents by relying on models of other agents to predict their actions, and on optimal choices of agent’s own behaviors given these predictions. The two main elements, first of predicting actions of others, and second of choosing own action, are both handled from the standpoint of Bayesian decision theory. The descriptive power of decision-theoretic rationality, which describes actions of rational individuals, is used to predict actions of other agents. The prescriptive aspect of decision-theoretic rationality, which dictates that optimal actions chosen are the ones that maximize expected utility, is used for agent to select its own action.

2 Background: Partially Observable Markov Decision Processes

A partially observable Markov decision processes (POMDP) [6, 13, 16, 20] of an agent i is defined as

$$POMDP_i = \langle S, A_i, T_i, \Omega_i, O_i, R_i \rangle \quad (1)$$

where: S is a set of possible states of the environment (defined as the reality external to the agent i), A_i is a set of actions agent i can execute, T_i is a transition function – $T_i : S \times A_i \times S \rightarrow [0, 1]$ which describes results of agent i ’s actions, Ω_i is the set of observations that the agent i can make, O_i is the agent’s observation function – $O_i : \Omega_i \times S \times A_i \rightarrow [0, 1]$ which specifies probabilities of observations if agent executes various actions that result in different states, R_i is the reward function representing the agent i ’s preferences – $R_i : S \times A_i \rightarrow \mathbf{R}$.

The belief update and the optimal solutions (depending on an optimality criterion) to POMDPs are described in the literature [6, 13, 16, 20].

2.1 Optimality Criteria

The agent’s optimality criterion, OC_i , is needed to specify how rewards acquired over time are handled. Commonly used criteria include:

- A finite horizon criterion, in which the agent maximizes the expected value of the sum of first T rewards: $E(\sum_{t=0}^T r_t)$, where r_t is a reward obtained at time t and T is the length of the horizon. We will denote this criterion as fh^T .

- An infinite horizon criterion with discounting, according to which the agent maximizes $E[\sum_{t=0}^{\infty} \gamma^t r_t]$, where $0 < \gamma < 1$ is a discount factor. We will denote this criterion as ih^γ .
- An infinite horizon criterion with averaging, according to which the agent maximizes the average reward per time step. We will denote this as ih^{AV} .

In what follows, we concentrate on infinite horizon criterion with discounting, but our approach can be easily adapted to other criteria.

2.2 Agent Types and Frames

The POMDP definition above includes parameters that permit us to compute an agent's optimal behavior, conditioned on its beliefs. Let us collect these implementation independent factors into a construct we call an agent i 's type.

Definition 1. (Type) A type of an agent i is $\theta_i = \langle b_i, A_i, \Omega_i, T_i, O_i, R_i, OC_i \rangle$, where b_i is agent i 's state of belief (an element of $\Delta(S)$), OC_i is its optimality criterion, and the rest of the elements are as defined before. Let Θ_i be the set of agent i 's types.

Given type, θ_i , and under the assumption that the agent is Bayesian-rational, the set of agent's optimal actions will be denoted as $OPT(\theta_i)$. In the next section, we generalize the notion of type to situations which include interactions with other agents; it then coincides with the notion of type used in Bayesian games [12, 10].

It is convenient to define the notion of a *frame*, $\hat{\theta}_i$, of agent i :

Definition 2. (Frame) A frame of an agent i is $\hat{\theta}_i = \langle A_i, \Omega_i, T_i, O_i, R_i, OC_i \rangle$. Let $\hat{\Theta}_i$ be the set of agent i 's frames.

For brevity one can write a type as consisting of an agent's belief together with its frame: $\theta = \langle b_i, \hat{\theta}_i \rangle$.

3 Interactive POMDPs

As we mentioned, our intention is to generalize POMDPs to handle presence of other agents. We do this by including descriptions of other agents (their types, for example), as well as physical aspects of agent's own description (i.e., their own frames) in the state space. For simplicity of presentation we consider an agent i that is interacting with one other agent j .

Definition 3. (I-POMDP) An interactive POMDP of agent i , $I\text{-POMDP}_i$, is:

$$I\text{-POMDP}_i = \langle IS_i, A, T_i, \Omega_i, O_i, R_i \rangle \quad (2)$$

where:

- IS_i is a set of **interactive** states defined as $IS_i = S \times M_j$ where S is the set of states of the physical environment, and M_j is the set of possible models of agent j . Models of other agents are included in the state space to allow an agent to have beliefs over them.

Models of agents include factors relevant to the agents' behavior. Analogously to states of the world, models of agents are intended to be rich enough to allow informed prediction about behavior. Agent i maintains its belief about the interactive state as a probability distribution over IS_i . Let us

note that the states are subjective; the reality external to each agent, say i , is different in that it includes agents other than i – in this case agent j .

A *general* model of an agent, j , is a function $m_j : H_j \rightarrow \Delta(A_j)$, i.e., a mapping from j 's observable histories to probabilistic predictions of j 's behavior. Let M_j be the set of all models of j that are computable. One example of a model is obtained during the fictitious play considered in game theory [9, 22]. In *fictitious play* model probabilities of agent's future actions are estimated as frequencies of actions observed in the past. Another simple version is a *no-information* model [11], according to which actions are independent of history and are assumed to occur with uniform probabilities, $1/|A_j|$, each.

Perhaps the most interesting model is the *intentional* model, defined to be the agent j 's type, $\theta_j = \langle b_j, A, \Omega_j, T_j, O_j, R_j, OC_j \rangle$, together with the assumption that agent j is Bayesian rational. Agent j 's belief is a probability distribution over the states of the world, the models of the agent i , and frames of itself – $b_j \in \Delta(S \times M_i)$. The notion of an intentional model, or type, we use here coincides with the notion of type in game theory, where it is defined as consisting of all of the agent i 's private information relevant to its decision making [12, 10]. In particular, if agents' beliefs are private information, then their types involve possibly infinitely nested beliefs over others' types and their beliefs about others. They have been called knowledge-belief structures and type hierarchies in game theory [1, 2, 7, 19], and are related to recursive model structures in our prior work [11]¹.

- $A = A_i \times A_j$ is the set of joint moves of all agents.

- T_i is a transition function $T_i : IS_i \times A \times IS_i \rightarrow [0, 1]$ which describes results of agents' actions. Actions can change the physical state, as well as the models of other agents and the agent's own frame, for example by changing the observation function of one or both agents. One can model communicative actions as changing the beliefs of the agents directly, but it may be more appropriate to model communication as action that can be observed by others and thus can change their beliefs indirectly.

One can make the following assumption about T_i :

Definition 4. (Belief Non-manipulability (BNM)) *Agents' actions do not change the agents' beliefs directly. Formally, for all $s, b_j, \hat{\theta}_j, s', b'_j, \hat{\theta}'_j$ we have:*

$$T_i((s, \langle b_j, \hat{\theta}_j \rangle), a, (s', \langle b'_j, \hat{\theta}'_j \rangle)) > 0 \text{ only if } b_j = b'_j. \quad (3)$$

Belief non-manipulability is justified in usual cases of interacting autonomous agents. Autonomy precludes direct “mind control” and implies that agents' belief states can be changed only indirectly, typically by changing the environment in a way observable to them.² As we mentioned the agent's actions are still allowed to change, say, their observation capabilities and their preferences.³

- Ω_i is defined as before in POMDP model.

- O_i is an observation function $O_i : \Omega_i \times IS_i \times A \rightarrow [0, 1]$.

One can make the following assumption about the observation function.

¹Implicit in the definition of nested beliefs is the assumption of coherency [7].

²In other words, agents' beliefs do change, just like in POMDPs, but as a result of belief update after an observation, not as a direct result of any of the agents' actions.

³One can strengthen the notion of autonomous agents by postulating that their preferences are non-manipulable as well, but we do not go into this here for simplicity.

Definition 5. (Belief Non-observability (BNO)) Agents cannot observe other’s beliefs directly. Formally, for all $o, s, b_j, \hat{\theta}_j$, and b'_j , we have:

$$O_i(o, (s, \langle b_j, \hat{\theta}_j \rangle), a) = O_i(o, (s, \langle b'_j, \hat{\theta}_j \rangle), a). \quad (4)$$

The BNO assumption does not imply that the other elements of the agent’s type are fully observable. For example, it is unlikely that an agent’s reward function would be directly observable by other agents, but we do not get into this issue here for simplicity.

- R_i is defined as $R_i : IS_i \times A \rightarrow \mathbf{R}$. We allow the agent to have preferences over physical states and models of all agents, but usually only the physical state will matter.

3.1 Belief Update in *I-POMDPs*

We will show that, as in POMDPs, agent’s beliefs over their interactive states are *sufficient statistics*, i.e., they fully summarize the agent’s observable histories. Further, we need to show how beliefs are updated after agent’s action and observation, and how solution is defined. There are two differences that complicate belief update, when compared to POMDPs. First, since the state of the physical environment depends on the actions performed by both agents the prediction of how the physical state changes has to be made based on the predicted actions of the other agent. The probabilities of other’s actions are obtained based on their models. Thus, unlike in Bayesian and stochastic games, we do not assume that actions are fully observable by other agents. Rather, agents can attempt to infer what actions other agents have performed by sensing their results on the environment.

Second, changes in the models of the other agents have to be included in the update. Some of these changes may be directly attributed to the agents’ actions,⁴ but, more importantly, the update of the other agent’s beliefs due to its new observation has to be included. In other words, the agent has to update its beliefs based on what it anticipates that the other agent observes and how it updates. As could be expected, the update of the possibly infinitely nested belief over other’s types is, in general, only asymptotically computable.

Proposition 1. (Sufficiency) In an interactive POMDP of agent i , i ’s current belief, i.e., the probability distribution over the set $S \times M_j$, is a sufficient statistic for the past history of i ’s observations.

The next proposition defines the agent i ’s belief update function, $b'_i(is^t) = Pr(is^t | o_i^t, a_i^{t-1}, b_i^{t-1})$, where $is^t \in IS_i$ is an interactive state. We will use the belief state estimation function, SE_{θ_i} , as an abbreviation for belief updates for individual states so that $b'_i = SE_{\theta_i}(b_i^{t-1}, a_i^{t-1}, o_i^t)$. $\tau_{\theta_i}(b_i^{t-1}, a_i^{t-1}, o_i^t, b_i^t)$ will stand for $Pr(b_i^t | b_i^{t-1}, a_i^{t-1}, o_i^t)$. We will also define the set of type-dependent optimal actions of an agent, $OPT(\theta_i)$.

Proposition 2. (Belief Update) Under the BNM and BNO assumptions, the belief update function for interactive POMDP $\langle IS_i, A, T_i, \Omega_i, O_i, R_i \rangle$ is:

$$\begin{aligned} b'_i(is^t) &= \beta \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} Pr(a_j^{t-1} | m_j^{t-1}) O_i(is^t, a^{t-1}, o_i^t) \\ &\times \sum_{o_j^t} \tau_{\theta_j}(b_j^{t-1}, a_j^{t-1}, o_j^t, b_j^t) T_i(is^{t-1}, a^{t-1}, is^t) \\ &\times O_j(is_j^t, a^{t-1}, o_j^t) \end{aligned} \quad (5)$$

where $is = (s, m_j)$, $is_j = (s, m_i)$, b_j^{t-1} and b_j^t are the belief elements of m_j^{t-1} and m_j^t if these models are intentional, respectively, β is a normalizing constant, O_j is the observation function in m_j^t

⁴For example, actions may change the agents’ observation capabilities directly.

(assuming it is intentional model), and $Pr(a_j^{t-1}|m_j^{t-1})$ is the probability of other agent's action given its model, i.e., the probability that a_j^{t-1} is Bayesian rational for that type of agent. If the model is the other agent's type, θ_j , then this probability is equal to $\frac{1}{|OPT(\theta_j)|}$ if $a_j^{t-1} \in OPT(\theta_j)$, and it is equal to zero otherwise. We define OPT below. If the model of agent j is not intentional then the probability of actions are given directly by the model and the summation over o_j, τ_{θ_j} , and the last line drop out.

Proof of Propositions 1 and 2. We start with Proposition 2, by applying the Bayes Theorem:

$$\begin{aligned}
b_i^t(is^t) &= Pr(is^t|o_i^t, a_i^{t-1}, b_i^{t-1}) = \frac{Pr(is^t, o_i^t, a_i^{t-1}, b_i^{t-1})}{Pr(o_i^t, a_i^{t-1}, b_i^{t-1})} = \frac{Pr(is^t, o_i^t|a_i^{t-1}, b_i^{t-1})Pr(a_i^{t-1}, b_i^{t-1})}{Pr(o_i^t|a_i^{t-1}, b_i^{t-1})Pr(a_i^{t-1}, b_i^{t-1})} \\
&= \frac{Pr(is^t, o_i^t|a_i^{t-1}, b_i^{t-1})}{Pr(o_i^t|a_i^{t-1}, b_i^{t-1})} = \beta \sum_{is^{t-1}} Pr(is^t, o_i^t|a_i^{t-1}, is^{t-1})b_i^{t-1}(is^{t-1}) \\
&= \beta \sum_{is^{t-1}} \sum_{a_j^{t-1}} Pr(is^t, o_i^t|a_i^{t-1}, a_j^{t-1}, is^{t-1})Pr(a_j^{t-1}|a_i^{t-1}, is^{t-1})b_i^{t-1}(is^{t-1}) \\
&= \beta \sum_{is^{t-1}} \sum_{a_j^{t-1}} Pr(is^t, o_i^t|a^{t-1}, is^{t-1})Pr(a_j^{t-1}|is^{t-1})b_i^{t-1}(is^{t-1}) \\
&= \beta \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} Pr(a_j^{t-1}|m_j^{t-1})Pr(o_i^t|is^t, a^{t-1}, is^{t-1})Pr(is^t|a^{t-1}, is^{t-1}) \\
&= \beta \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} Pr(a_j^{t-1}|m_j^{t-1})Pr(o_i^t|is^t, a^{t-1})Pr(is^t|a^{t-1}, is^{t-1}) \\
&= \beta \sum_{is^{t-1}} b_i^{t-1}(is^{t-1}) \sum_{a_j^{t-1}} Pr(a_j^{t-1}|m_j^{t-1})O_i(is^t, a^{t-1}, o_i^t)Pr(is^t|a^{t-1}, is^{t-1})
\end{aligned} \tag{6}$$

$$Pr(is^t|a^{t-1}, is^{t-1}) = \sum_{o_j^t} Pr(is^t|a^{t-1}, is^{t-1}, o_j^t)Pr(o_j^t|a^{t-1}, is^{t-1}) \tag{7}$$

In order to simplify the term $Pr(is^t|a^{t-1}, is^{t-1}, o_j^t)$, let us substitute the interactive state is^t with its components, $is^t = (s^t, m_j^t)$, and if m_j^t is intentional, with $(s^t, b_j^t, \hat{\theta}_j^t)$.

$$\begin{aligned}
Pr(is^t|a^{t-1}, is^{t-1}, o_j^t) &= Pr(s^t, b_j^t, \hat{\theta}_j^t|a^{t-1}, is^{t-1}, o_j^t) \\
&= Pr(b_j^t|s^t, \hat{\theta}_j^t, a^{t-1}, is^{t-1}, o_j^t)Pr(s^t, \hat{\theta}_j^t|a^{t-1}, is^{t-1}, o_j^t)
\end{aligned}$$

In the above equation, the first term on the right-hand side is 1 if Agent j 's belief update, $SE_{\theta_j}(b_j^{t-1}, a_j^{t-1}, o_j^t)$ generates a belief state equal to b_j^t . This term drops out if m_j^t is not an intentional model. The action pair a^{t-1} may change the physical state, Agent j 's, and Agent i 's model. The second term on the right-hand side captures this transition. We utilize the BNM assumption to replace the second term with the transition function.

$$Pr(is^t|a^{t-1}, is^{t-1}, o_j^t) = \tau_{\theta_j^t}(b_j^{t-1}, a_j^{t-1}, o_j^t, b_j^t)T_i(is^{t-1}, a^{t-1}, is^t)$$

In order to evaluate term 2 of equation Eq. 7, we introduce an intermediate state $is^{t-1/2}$. The intermediate state results after the action pair a^{t-1} but before the observations are perceived.

$$\begin{aligned}
Pr(o_j^t|a^{t-1}, is^{t-1}) &= \sum_{is^{t-1/2}} Pr(o_j^t|a^{t-1}, is^{t-1}, is^{t-1/2})Pr(is^{t-1/2}|a^{t-1}, is^{t-1}) \\
&= \sum_{is^{t-1/2}} Pr(o_j^t|a^{t-1}, is^{t-1/2})Pr(is^{t-1/2}|a^{t-1}, is^{t-1})
\end{aligned}$$

In the first term of the above equation, the BNO assumption makes it possible to replace $is^{t-1/2}$ with is^t .

$$\begin{aligned}
Pr(o_j^t|a^{t-1}, is^{t-1}) &= \sum_{is^{t-1/2}} Pr(o_j^t|a^{t-1}, is^t)Pr(is^{t-1/2}|a^{t-1}, is^{t-1}) \\
&= Pr(o_j^t|a^{t-1}, is^t) \sum_{is^{t-1/2}} Pr(is^{t-1/2}|a^{t-1}, is^{t-1}) \\
&= Pr(o_j^t|a^{t-1}, is^t) \\
&= O_j(is_j^t, a^{t-1}, o_j^t)
\end{aligned}$$

We now replace the summand of Eq. 7 with the evaluated expressions.

$$\begin{aligned} Pr(is^t|a^{t-1}, is^{t-1}) &= \sum_{o_j^t} \tau_{\theta_j^t}(b_j^{t-1}, a_j^{t-1}, o_j^t, b_j^t) T_i(is^{t-1}, a^{t-1}, is^t) \\ &\times O_j(is_j^t, a^{t-1}, o_j^t) \end{aligned} \quad (8)$$

The final equation for our belief update (Eq. 5) results by replacing Eq. 8 into Eq. 6.

Since Proposition 2 expresses the belief $b_i^t(is^t)$ in terms of parameters in the previous time step only, the Proposition 1 holds as well. \square

Intuitively, Proposition 1 holds since, as in POMDPs [29], all of the dynamic elements (state transitions and observations) of the model depend on the previous state, not on prior observations and actions.

Proposition 1 and Eq. 5 have a lot in common with belief update in POMDPs, as should be expected. Both depend on agent i 's observation and transition functions. However, since agent i 's observations also depend on agent j 's actions, the probabilities of various actions of j have to be included (in the first line of Eq. 5.) Further, since the update of agent j 's model depends on what j observes, the probabilities of various observations have to be included (in the last line of Eq. 5.) The update of j 's beliefs is included by adding the τ_{θ_j} term.

If none of the models m_j are intentional the belief update in I-POMDPs reduces to form similar to POMDPs. If some models are intentional the belief could be infinitely nested and the belief update can be calculated only asymptotically. The belief update can easily be generalized to the setting where more than one other agents co-exist with agent i .

3.2 Value Function and Solutions in I-POMDPs

Analogously to POMDPs, each belief state in I-POMDP has an associated value reflecting the maximum payoff the agent can expect in this belief state:

$$U(\theta_i) = \max_{a_i \in A_i} \sum_{is} b_i(is) ER_i^{a_i}(is, a_i) + \gamma \sum_{o_i \in \Omega_i} Pr(o_i|a_i, b_i) U(SE_{\theta_i}(b_i, a_i, o_i)) \quad (9)$$

where, $ER_i^{a_i}(is, a_i) = \sum_{a_j} R_i(is, a_i, a_j) Pr(a_j|m_j)$ (since $is = (s, m_j)$).

Agent i 's optimal action, a^* , for the case of infinite horizon criterion with discounting, is an element of the set of optimal actions for the belief state, $OPT(\theta_i)$, which is defined as:

$$OPT(\theta_i) = \operatorname{argmax}_{a_i \in A_i} \sum_{is} b_i(is) ER_i^{a_i}(is, a_i) + \gamma \sum_{o_i \in \Omega_i} Pr(o_i|a_i, b_i) U(SE_{\theta_i}(b_i, a_i, o_i)) \quad (10)$$

Since the beliefs could be infinitely nested approximations that involve terminating the nesting of beliefs, for example with a no-information model, involve solving a finite number of traditional POMDPs, and their complexity is PSPACE-hard [26]. Including more information residing in deeper levels of nesting results in better approximations. We investigated error bounds of such approximations experimentally in [23] in myopic settings; formal proof of the error bounds for longer time horizons remain subject of future work.

4 Conclusions

This paper proposes a decision-theoretic approach to game theory as a paradigm for designing agents that are able to intelligently interact and coordinate actions with other agents in multi-agent environments. We defined a general multi-agent version of partially observable Markov decision processes, called interactive POMDP's, and illustrated assumptions, some basic properties and solution method.

This line of work opens a wide area of fertile future research that integrates frameworks for sequential planning, like POMDPs, with elements of game theory and inductive learning.

References

- [1] Robert J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.
- [2] Robert J. Aumann and Aviad Heifetz. *Handbook of Game Theory with Economic Applications*, volume 3. Elsevier Science, 2002.
- [3] Daniel S. Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of Operations Research*, 2002.
- [4] Ken Binmore. *Essays on Foundations of Game Theory*. Pittman, 1982.
- [5] Craig Boutilier. Sequential optimality and coordination in multiagent systems. In *Sixteenth International Joint Conference on Artificial Intelligence*, pages 478–485, 1999.
- [6] Craig Boutilier, Thomas Dean, and Steve Hanks. Handbook of game theory with economic applications. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.
- [7] Adam Brandenburger and Eddie Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59:189–198, 1993.
- [8] Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
- [9] Drew Fudenberg and David K Levine. *The Theory of Learning in Games*. MIT Press, 1998.
- [10] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.
- [11] Piotr Gmytrasiewicz and Edmund Durfee. Rational coordination in multi-agent environments. *Autonomous Agents and Multiagent Systems Journal*, 3(4):319–350, 2000.
- [12] John C. Harsanyi. Games with incomplete information played by 'bayesian' players. *Management Science*, 14(3):159–182, 1967.
- [13] Milos Hauskrecht. Value-function approximations for partially observable markov decision process. *Journal of Artificial Intelligence*, 13:33–94, 2000.
- [14] J. Hu and M. P. Wellman. Multiagent reinforcement learning: Theoretical framework and an algorithm. In *Fifteenth International Conference on Machine Learning*, pages 242–250, 1998.
- [15] Joseph Kadane and Patrick Larkey. Subjective probability and the theory of games. *Management Science*, 28(2):113–120, 1982.
- [16] Leslie Kaelbling, Michael Littman, and Anthony Cassandra. Planning and acting partially observable stochastic domains. *Artificial Intelligence*, 2, 1998.
- [17] Daphne Koller and Brian Milch. Multi-agent influence diagrams for representing and solving games. In *Seventeenth International Joint Conference on Artificial Intelligence*, pages 1027–1034, August 2001.
- [18] Michael Littman. Markov games as a framework for multiagent reinforcement learning. In *International Conference on Machine Learning*, 1994.
- [19] J.F. Mertens and S. Zamir. Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.
- [20] George Monahan. A survey of partially observable markov decision processes. theory, models and algorithms. *Management Science*, pages 1–16, 1982.
- [21] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

- [22] J. Nachbar. Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, 19:59–89, 1990.
- [23] Sanguk Noh and Piotr Gmytrasiewicz. Identifying the scope of modeling for time-critical multiagent decision-making. In *17th International Conference on Artificial Intelligence*, pages 1043–1048, August 2001.
- [24] J. M. Ooi and G.W.Wornell. Decentralized control of a multiple broadcast channel. In *35th Conference on Decision and Control*, 1996.
- [25] Guillermo Owen. *Game Theory: Second Edition*. Academic Press, 1982.
- [26] C.H. Papadimitriou and J.N. Tsitsiklis. The complexity of markov decision processes. *Mathematical Journal of Operations Research*, 12(3):441–450, 1987.
- [27] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (Second Edition)*. Prentice Hall, 2003.
- [28] Robert Samuel Simon. Locally finite knowledge structures. Technical Report 275, Center for Rationality and Interactive Decision Theory, Hebrew University, Jerusalem, Israel, October 2001.
- [29] Richard Smallwood and Edward Sondik. The optimal control of partially observable markov decision processes over a finite horizon. *Operations Research*, 21:1071–1088, 1973.
- [30] Milind Tambe, Ranjit Nair, David Pynadath, and Stacy Marsella. Towards computing optimal policies for decentralized pomdps. In *Notes of the 2002 AAAI Workshop on Game Theoretic and Decision Theoretic Agents*, August 2002.