Limited Lookahead in Imperfect-Information Games

Christian Kroer and Tuomas Sandholm
Computer Science Department
Carnegie Mellon University
ckroer@cs.cmu.edu, sandholm@cs.cmu.edu

Abstract

Limited lookahead has been studied for decades in perfect-information games. This paper initiates a new direction via two simultaneous deviation points: generalization to imperfect-information games and a game-theoretic approach. The question of how one should act when facing an opponent whose lookahead is limited is studied along multiple axes: lookahead depth, whether the opponent(s), too, have imperfect information, and how they break ties. We characterize the hardness of finding a Nash equilibrium or an optimal commitment strategy for either player, showing that in some of these variations the problem can be solved in polynomial time while in others it is PPAD-hard or NP-hard. We proceed to design algorithms for computing optimal commitment strategies for when the opponent breaks ties 1) favorably, 2) according to a fixed rule, or 3) adversarially. The impact of limited lookahead is then investigated experimentally. The limited-lookahead player often obtains the value of the game if she knows the expected values of nodes in the game tree for some equilibrium, but we prove this is not sufficient in general. Finally, we study the impact of noise in those estimates and different lookahead depths. This uncovers a lookahead pathology.

1 Introduction

Limited lookahead has been a central topic in AI game playing for decades. To date, it has been studied in single-agent settings and perfect-information games—specifically in well-known games such as chess, checkers, Go, etc., as well as in random game tree models [Korf, 1990; Pearl, 1981; Pearl, 1983; Nau, 1983; Nau et al., 2010; Bouzy and Cazenave, 2001; Ramanujan, Sabharwal, and Selman, 2010; Ramanujan and Selman, 2011]. In this paper, we initiate the game-theoretic study of limited lookahead in imperfect-information games. Such games are more broadly applicable to practical settings—for example auctions, negotiations, security, cybersecurity, and medical settings—than perfect-information games. Mirrokni, Thain, and Vetta [2012] conducted a game-theoretic analysis of lookahead, but they consider only perfect-information games, and the results are for four specific games rather than broad classes of games. Instead, we analyze the questions for imperfect information and general-sum extensive-form games.

As is typical in the literature on limited lookahead in perfect-information games, we derive our results for a two-agent setting. One agent is a rational player (Player r) trying to optimally exploit a limited-lookahead player (Player l). Our results extend immediately to one rational player and more than one limited-lookahead player, as long as the latter all break ties according to the same scheme (statically, favorably, or adversarially—as described later in the paper). This is because such a group of limited-lookahead players can be treated as one from the perspective of our results.

The type of limited-lookahead player we introduce is analogous to that in the literature on perfect-information games. Specifically, we let the limited-lookahead player l have a node evaluation function h that places numerical values on all nodes in the game tree. Given a strategy for the rational player, at each information set at some depth i, Player l picks an action that maximizes the expected value of the evaluation function at depth $i + k$, assuming optimal play between those levels. Our study is the game-theoretic, imperfect-information generalization of lookahead questions studied in the literature and therefore interesting in its own right. The model also has applications such as biological games, where the goal is to steer an evolution or adaptation process (which typically acts myopically with lookahead 1) [Sandholm, 2015] and security games where opponents are often assumed to be myopic (as makes sense when the number of adversaries is large [Yin et al., 2012]). Furthermore, investigating how well a rational player can exploit a limited-lookahead player lends insight into the limitations of using limited-lookahead algorithms in multiagent decision making.

We then design algorithms for finding an optimal strategy to commit to for the rational player. We focus on this rather than equilibrium computation because the latter seems nonsensical in this setting: the limited-lookahead player determining a Nash equilibrium strategy would require her to reason about the whole game for the rational player's strategy, which rings contrary to the limited-lookahead assumption. Computing optimal strategies to commit to in standard rational settings has previously been studied in normal-form games [Conitzer and Sandholm, 2006] and extensive-form games [Letchford and Conitzer, 2010], the latter implying some complexity results for our setting as we will discuss.

As in the literature on lookahead in perfect-information games, a potential weakness of our approach is that we require knowing the evaluation function h (but make no other assumptions about what information h encodes). In practice, this function may not be known. As in the perfect-information setting, this can lead to the rational exploiter being exploited.

2 Extensive-form games

We start by defining the class of games that the players will play, without reference to limited lookahead. An *extensive-form game* Γ is a tuple $\langle N, A, S, Z, \mathcal{H}, \sigma_0, u, \mathcal{I}\rangle$. N is the set of players. A is the set of all actions in the game. S is a set of nodes corresponding to sequences of actions. They describe a tree with root node $s^r \in S$. At each node s, it is the turn of some Player i to move. Player i chooses among actions A_s, and each branch at s denotes a different choice in A_s. Let t^s_a be the node transitioned to by taking action $a \in A_s$ at node s. The set of all nodes where Player i is active is called S_i. $Z \subset S$ is the set of leaf nodes. The utility function of Player i is $u_i : Z \to \mathbb{R}$, where $u_i(z)$ is the utility to Player i when reaching node z. We assume, without loss of generality, that all utilities are non-negative. Z_s is the subset of leaf nodes reachable from a node s. $\mathcal{H}_i \subseteq \mathcal{H}$ is the set of heights in the game tree where Player i acts. Certain nodes correspond to stochastic outcomes with a fixed probability distribution. Rather than treat those specially, we let Nature be a static player acting at those nodes. $\mathcal{H}_0$ is the set of heights where Nature acts. σ_0 specifies the probability distribution for Nature, with $\sigma_0(s, a)$ denoting the probability of Nature choosing outcome a at node s. Imperfect information is represented in the game model using information sets. $\mathcal{I}_i \subseteq \mathcal{I}$ is the set of information sets where Player i acts. $\mathcal{I}_i$ partitions S_i. For nodes $s_1, s_2 \in I, I \in \mathcal{I}_i$, Player i cannot distinguish among them, and so $A_{s_1} = A_{s_2} = A_I$.

We denote by $\sigma_i : S_i \to [0, 1]$ a *behavioral strategy* for Player i. For each information set $I \in \mathcal{I}_i$, it assigns a probability distribution over A_I, the actions at the information set. $\sigma_i(I, a)$ is the probability of playing action a at information set I. A *strategy profile* $\sigma = (\sigma_0, \ldots, \sigma_n)$ consists of a behavioral strategy for each player. We will often use $\sigma(I, a)$ to mean $\sigma_i(I, a)$, since the information set specifies which Player i is active. As described above, randomness external to the players is captured by the Nature outcomes σ_0.

Let the probability of going from node s to node $\hat{s}$ under strategy profile σ be $\pi^\sigma(s, \hat{s}) = \Pi_{\langle \bar{s}, \bar{a}\rangle \in X_{s,\hat{s}}} \sigma(\bar{s}, \bar{a})$ where $X_{s,\hat{s}}$ is the set of node-action pairs on the path from s to $\hat{s}$. We let the probability of reaching node s be $\pi^\sigma(s) = \pi^\sigma(s^r, s)$, the probability of going from the root node to s. Let $\pi^\sigma(I) = \sum_{s\in I} \pi^\sigma(s)$ be the probability of reaching any node in I. Due to perfect recall, we have $\pi^\sigma_i(I) = \pi^\sigma_i(s)$ for all $s \in I$. For probabilities over Nature, $\pi^\sigma_0 = \pi^{\bar{\sigma}}_0$ for all $\sigma, \bar{\sigma}$, so we can ignore the strategy profile superscript and write π_0. Finally, for all behavioral strategies, the subscript $-i$ refers to the same definition, excluding Player i. For example, $\pi^\sigma_{-i}(s)$ denotes the probability of reaching s over the actions of the players other than i, that is, if i played to reach s with probability 1.

3 Model of limited lookahead

We now describe our model of limited lookahead. We use the term optimal *hypothetical* play to refer to the way the limited-lookahead agent thinks she will play when looking ahead from a given information set. In actual play part way down that plan, she may change her mind because she will then be able to see to a deeper level of the game tree.

Let k be the lookahead of Player l, and $S^k_{I,a}$ the nodes at lookahead depth k below information set I that are reachable (through some path) by action a. As in prior work in the perfect-information game setting, Player l has a node-evaluation function $h : S \to \mathbb{R}$ that assigns a heuristic numerical value to each node in the game tree.

Given a strategy σ_r for the other player and fixed action probabilities for Nature, Player l chooses, at any given information set $I \in \mathcal{I}_l$ at depth i, a (possibly mixed) strategy whose support is contained in the set of actions that maximize the expected value of the heuristic function at depth $i + k$, assuming optimal hypothetical play by her ($\max_{\sigma_l}$ in the formula below). We will denote this set by $A^*_I =$

$$\{a : a \in \arg\max_{a\in A_I} \max_{\sigma_l} \sum_{s\in I} \frac{\pi^{\sigma_{-l}}(s)}{\pi^{\sigma_{-l}}(I)} \sum_{s'\in S^k_{I,a}} \pi^\sigma(t^s_a, s')h(s')\},$$

where $\sigma = \{\sigma_l, \sigma_r\}$ is the strategy profile for the two players. Here moves by Nature are also counted toward the depth of the lookahead. The model is flexible as to how the rational player chooses σ_r and how the limited-lookahead player chooses a (possibly mixed) strategy with supports within the sets A^*_I. For one, we can have these choices be made for both players simultaneously according to the Nash equilibrium solution concept. As another example, we can ask how the players should make those choices if one of the players gets to make, and commit to, all her choices before the other.

4 Complexity

In this section we analyze the complexity of finding strategies according to these solution concepts.

Nash equilibrium. Finding a Nash equilibrium when Player l either has information sets containing more than one node, or has lookahead at least 2, is PPAD-hard [Papadimitriou, 1994]. This is because finding a Nash equilibrium in a 2-player general-sum normal-form game is PPAD-hard [Chen, Deng, and Teng, 2009], and any such game can be converted to a depth 2 extensive-form game, where the general-sum payoffs are the evaluation function values.

If the limited-lookahead player only has singleton information sets and lookahead 1, an optimal strategy can be trivially computed in polynomial time in the size of the game tree for the limited-lookahead player (without even knowing the other player's strategy σ_r): for each of her information sets, we simply pick an action that has highest immediate heuristic value. To get a Nash equilibrium, what remains to be done is to compute a best response for the rational player, which can also be easily done in polynomial time [Johanson et al., 2011].

Commitment strategies. Next we study the complexity of finding commitment strategies (that is, finding a strategy for the rational player to commit to, where the limited lookahead player then responds to that strategy.). The complexity depends on whether the game has imperfect information (information sets that include more than one node) for the limited-lookahead player, how far that player can look ahead, and how she breaks ties in her action selection.

When ties are broken adversarially, the choice of response depends on the choice of strategy for the rational player. If Player l has lookahead one and no information sets, it is easy to find the optimal commitment strategy: the set of optimal actions A^*_s for any node $s \in S_l$ can be precomputed, since Player r does not affect which actions are optimal. Player l will then choose actions from these sets to minimize the utility of Player r. We can view the restriction to a subset of actions as a new game, where Player l is a rational player in a zero-sum game. An optimal strategy for Player r to commit to is then a Nash equilibrium in this smaller game. This is solvable in polynomial time by an LP that is linear in the size of the game. The problem is hard without either of these assumptions. This is shown in an extended online version.

5 Algorithms

In this section we will develop an algorithm for solving the hard commitment-strategy case. Naturally its worst-case runtime is exponential. As mentioned in the introduction, we focus on commitment strategies rather than Nash equilibria because Player l playing a Nash equilibrium strategy would require that player to reason about the whole game for the opponent's strategy. Further, optimal strategies to commit to are desirable for applications such as biological games [Sandholm, 2015] (because evolution is responding to what we are doing) and security games [Yin et al., 2012] (where the defender typically commits to a strategy).

Since the limited-lookahead player breaks ties adversarially, we wish to compute a strategy that maximizes the worst-case best response by the limited-lookahead player. For argument's sake, say that we were given $\mathcal{A}$, which is a fixed set of pairs, one for each information set I of the limited-lookahead player, consisting of a set of optimal actions A^*_I and one strategy for hypothetical play σ^I_l at I. Formally, $\mathcal{A} = \bigcup_{I \in \mathcal{I}_l} \langle A^*_I, \sigma^I_l \rangle$. To make these actions optimal for Player l, Player r must choose a strategy such that all actions in $\mathcal{A}$ are best responses according to the evaluation function of Player l. Formally, for all action triples $a, a^* \in \mathcal{A}, a' \notin \mathcal{A}$ (letting $\pi(s)$ denote probabilities induced by σ^I_l for the hypothetical play between I, a and s):

$$\sum_{s \in S^k_{I,a}} \pi(s) \cdot h(s) > \sum_{s \in S^k_{I,a'}} \pi(s) \cdot h(s) \tag{1}$$

$$\sum_{s \in S^k_{I,a}} \pi(s) \cdot h(s) = \sum_{s \in S^k_{I,a^*}} \pi(s) \cdot h(s) \tag{2}$$

Player r chooses a worst-case utility-maximizing strategy that satisfies (1) and (2), and Player l has to compute a (possibly mixed) strategy from $\mathcal{A}$ such that the utility of Player r is minimized. This can be solved by a linear program:

Theorem 1. *For some fixed choice of actions $\mathcal{A}$, Nash equilibria of the induced game can be computed in polynomial time by a linear program that has size $O(|S|) + O(\sum_{I \in \mathcal{I}_l} |A_I| \cdot \max_{s \in S} |A_s|^k)$.*

To prove this theorem, we first design a series of linear programs for computing best responses for the two players. We will then use duality to prove the theorem statement.

In the following, it will be convenient to change to matrix-vector notation, analogous to that of von Stengel [1996], with some extensions. Let $A = -B$ be matrices describing the utility function for Player r and the adversarial tie-breaking of Player l over $\mathcal{A}$, respectively. Rows are indexed by Player r sequences, and columns by Player l sequences. For sequence form vectors x, y, the objectives to be maximized for the players are then xAy, xBy, respectively. Matrices E, F are used to describe the sequence form constraints for Player r and l, respectively. Rows correspond to information sets, and columns correspond to sequences. Letting e, f be standard unit vectors of length $|\mathcal{I}_r|, |\mathcal{I}_l|$, respectively, the constraints $Ex = e, Fy = f$ describe the sequence form constraint for the respective players. Given a strategy x for Player r satisfying (1) and (2) for some $\mathcal{A}$, the optimization problem for Player l becomes choosing a vector of y' representing probabilities for all sequences in $\mathcal{A}$ that minimize the utility of Player r. Letting a prime superscript denote the restriction of each matrix and vector to sequences in $\mathcal{A}$, this gives the following primal (3) and dual (4) LPs:

$$\begin{aligned} \max_{y'} \quad & (x^T B')y' \\ & F'y' = f' \\ & y \geq 0 \end{aligned} \tag{3}$$

$$\begin{aligned} \min_{q'} \quad & q'^T f' \\ & q'^T F' \geq x^T B' \end{aligned} \tag{4}$$

where q' is a vector with $|\mathcal{A}| + 1$ dual variables. Given some strategy y' for Player l, Player r maximizes utility among strategies that induce $\mathcal{A}$. This gives the following best-response LP for Player r:

$$\begin{aligned} \max_{x} \quad & x^T(Ay') \\ & x^T E^T = e^T \\ & x \geq 0 \\ & x^T H_{\neg\mathcal{A}} - x^T H_{\mathcal{A}} \leq -\epsilon \\ & x^T G_{\mathcal{A}^*} = x^T G_{\mathcal{A}} \end{aligned} \tag{5}$$

where the last two constraints encode (1) and (2), respectively. The dual problem uses the unconstrained vectors p, v and constrained vector u and looks as follows

$$\begin{aligned} \min_{p,u,v} \quad & e^T p - \epsilon \cdot u \\ E^T p + (H_{\neg\mathcal{A}} - H_{\mathcal{A}})u + (G_{\mathcal{A}^*} - G_{\mathcal{A}})v &\geq A'y' \\ u &\geq 0 \end{aligned} \tag{6}$$

We can now merge the dual (4) with the constraints from the primal (5) to compute a minimax strategy: Player r chooses x, which she will choose to minimize the objective of (4),

$$\begin{aligned} \min_{x,q'} \quad & q'^T f' \\ q'^T F' - x^T B' &\geq 0 \\ -x^T E^T &= -e^T \\ x &\geq 0 \\ x^T H_{\mathcal{A}} - x^T H_{\neg\mathcal{A}} &\geq \epsilon \\ x^T G_{\mathcal{A}} - x^T G_{\mathcal{A}^*} &= 0 \end{aligned} \tag{7}$$

Taking the dual of this gives

$$\begin{aligned} \max_{y',p} \quad & -e^T p + \epsilon \cdot u \\ -E^T p + (H_{\mathcal{A}} - H_{\neg\mathcal{A}})u + (G_{\mathcal{A}} - G_{\mathcal{A}^*})v &\leq B'y' \\ F'y' &= f' \\ y, u &\geq 0 \end{aligned} \tag{8}$$

We are now ready to prove Theorem 1.

Proof. The LPs in Theorem 1 are (7) and (8). We will use duality to show that they provide optimal solutions to each of the best response LPs. Since $A = -B$, the first constraint in (8) can be multiplied by -1 to obtain the first constraint in (6) and the objective function can be transformed to that of (6) by making it a minimization. By the weak duality theorem, we get the following inequalities

$$q'^T f' \geq x^T B' y'; \text{ by LPs (3) and (4)}$$

$$e^T p - \epsilon \cdot u \geq x^T A' y'; \text{ by LPs (5) and (6)}$$

We can multiply the last inequality by -1 to get:

$$q'^T f' \geq x^T B' y' = -x^T A' y' \geq -e^T p + \epsilon \cdot u \quad (9)$$

By the strong duality theorem, for optimal solutions to LPs (7) and (8) we have equality in the objective functions $q'^T f' = -e^T p + \epsilon u$ which yields equality in (9), and thereby equality for the objective functions in LPs (3), (4) and for (5), (6). By strong duality, this implies that any primal solution x, q' and dual solution y', p to LPs (7) and (8) yields optimal solutions to the LPs (3) and (5). Both players are thus best responding to the strategy of the other agent, yielding a Nash equilibrium. Conversely, any Nash equilibrium gives optimal solutions x, y' for LPs (3) and (5). With corresponding dual solutions p, q', equality is achieved in (9), meaning that LPs (7) and (8) are solved optimally.

It remains to show the size bound for LP (7). Using sparse representation, the number of non-zero entries in the matrices A, B, E, F is linear in the size of the game tree. The constraint set $x^T H_{\mathcal{A}} - x^T H_{\neg\mathcal{A}} \geq \epsilon$, when naively implemented, is not. The value of a sequence $a \notin A_I^*$ is dependent on the choice among the cartesian product of choices at each information set I' encountered in hypothetical play below it. In practice we can avoid this by having a real-valued variable $v_I^d(I')$ representing the value of I' in lookahead from I, and constraints $v_I^d(I') \geq v_I^d(I', a)$ for each $a \in I'$, where $v_I^d(I', a)$ is a variable representing the value of taking a at I'. If there are more information sets below I' where Player l plays, before the lookahead depth is reached, we recursively constrain $v_I^d(I', a)$ to be:

$$v_I^d(I', a) \geq \sum_{\check{I} \in \mathcal{D}} v_I^d(\check{I}) \quad (10)$$

where $\mathcal{D}$ is the set of information sets at the next level where Player l plays. If there are no more information sets where Player l acts, then we constrain $v_I^d(I', a)$:

$$v_I^d(I', a) \geq \sum_{s \in S^k_{I', a}} \pi^\sigma_{-l} h(s) \quad (11)$$

Setting it to the probability-weighted heuristic value of the nodes reached below it. Using this, we can now write the constraint that a dominates all $a' \in I, a' \notin \mathcal{A}$ as:

$$\sum_{s \in S^k_{I,a}} \pi^\sigma(s) h(s) \geq v_I^d(I)$$

There can at most be $O(\sum_{I \in \mathcal{I}_l} |A_I|)$ actions to be made dominant. For each action at some information set I, there can be at most $O(\max_{s \in S} |A_s|^{\min\{k, k'\}})$ entries over all the constraints, where k' is the maximum depth of the subtrees rooted at I, since each node at the depth the player looks ahead to has its heuristic value added to at most one expression. For the constraint set $x^T G_{\mathcal{A}} - x^T G_{\mathcal{A}^*} = 0$, the choice of hypothetical plays has already been made for both expressions, and so we have the constraint

$$\sum_{s \in S^k_{I,a}} \pi^\sigma(s) h(s) = \sum_{s \in S^k_{I,a'}} \pi^{\sigma'}(s) h(s)$$

for all $I \in \mathcal{I}_l, a, a' \in I, \{a, \sigma^l\}, \{a', \sigma^{l,\prime}\} \in \mathcal{A}$, where

$$\sigma = \{\sigma_{-l}, \sigma^l\}, \sigma' = \{\sigma_{-l}, \sigma^{l,\prime}\}$$

There can at most be $\sum_{I \in \mathcal{I}_l} |A_I|^2$ such constraints, which is dominated by the size of the previous constraint set.

Summing up gives the desired bound. □

In reality we are not given $\mathcal{A}$. To find a commitment strategy for Player r, we could loop through all possible structures $\mathcal{A}$, solve LP (7) for each one, and select the one that gives the highest value. We now introduce a mixed-integer program (MIP) that picks the optimal induced game $\mathcal{A}$ while avoiding enumeration. The MIP is given in (12). We introduce Boolean sequence-form variables that denote making sequences suboptimal choices. These variables are then used to deactivate subsets of constraints, so that the MIP branches on formulations of LP (7), i.e., what goes into the structure $\mathcal{A}$. The size of the MIP is of the same order as that of LP (7).

$$\begin{aligned} \min_{x,q,z} \quad & q^T f \\ & q^T F \geq x^T B - zM \\ & Ex = e \\ & x^T H_{\mathcal{A}} \geq x^T H_{\neg\mathcal{A}} + \epsilon - (1-z)M \\ & x^T G_{\mathcal{A}} = x^T G_{\mathcal{A}^*} \pm (1-z)M \\ & \sum_{a \in A_I} z_a \geq z_{a'} \\ & x \geq 0, \quad z \in \{0,1\} \end{aligned} \quad (12)$$

The variable vector x contains the sequence form variables for Player r. The vector q is the set of dual variables for Player l. z is a vector of Boolean variables, one for each Player l sequence. Setting $z_a = 1$ denotes making the sequence a an inoptimal choice. The matrix M is a diagonal matrix with sufficiently large constants (e.g. the smallest value in B) such that setting $z_a = 1$ deactivates the corresponding constraint. Similar to the favorable-lookahead case, we introduce sequence form constraints $\sum_{a \in A_I} z_a \geq z_{a'}$ where a' is the parent sequence, to ensure that at least one action is picked when the parent sequence is active. We must also ensure that the incentivization constraints are only active for actions in $\mathcal{A}$:

$$x^T H_{\mathcal{A}} - x^T H_{\neg\mathcal{A}} \geq \epsilon - (1-z)M \quad (13)$$

$$x^T G_{\mathcal{A}} - x^T G_{\mathcal{A}^*} = 0 \pm (1-z)M$$

for diagonal matrices M with sufficiently large entries. Equality is implemented with a pair of inequality constraints $\{\leq, \geq\}$, where $\pm$ denotes adding or subtracting, respectively.

The values of each column constraint in (13) is implemented by a series of constraints. We add Boolean variables $\sigma_l^I(I', a')$ for each information set action pair I', a' that is potentially chosen in hypothetical play at I. Using our regular notation, for each a, a' where a is the action to be made dominant, the constraint is implemented by:

$$\sum_{s \in S^k_{I,a}} v^i(s) \geq v^d_I(I), \quad v^i(s) \leq \sigma^I_l(I', a') \cdot M \tag{14}$$

where the latter ensures that $v^i(s)$ is only non-zero if chosen in hypothetical play. We further need the constraint $v^i(s) \leq \pi^{\sigma}_{-l}(s)h(s)$ to ensure that $v^i(s)$, for a node s at the lookahead depth, is at most the heuristic value weighted by the probability of reaching s.

6 Experiments

In this section we experimentally investigate how much utility can be gained by optimally exploiting a limited-lookahead player. We conduct experiments on Kuhn poker [Kuhn, 1950], a canonical testbed for game-theoretic algorithms, and a larger simplified poker game that we call KJ. Kuhn poker consists of a three-card deck: king, queen, and jack. Each player antes 1. Each player is then dealt one of the three cards, and the third is put aside unseen. A single round of betting ($p = 1$) then occurs. In KJ, the deck consists of two kings and two jacks. Each player antes 1. A private card is dealt to each, followed by a betting round ($p = 2$), then a public card is dealt, follower by another betting round ($p = 4$). If no player has folded, a showdown occurs. For both games, each round of betting looks as follows:

- Player 1 can check or bet p.
  - If Player 1 checks Player 2 can check or raise p.
    - If Player 2 checks the betting round ends.
    - If Player 2 raises Player 1 can fold or call.
      - If Player 1 folds Player 2 takes the pot.
      - If Player 1 calls the betting round ends.
  - If Player 1 raises Player 2 can fold or call.
    - If Player 2 folds Player 1 takes the pot.
    - If Player 2 calls the betting round ends.

In Kuhn poker, the player with the higher card wins in a showdown. In KJ, showdowns have two possible outcomes: one player has a pair, or both players have the same private card. For the former, the player with the pair wins the pot. For the latter the pot is split. Kuhn poker has 55 nodes in the game tree and 13 sequences per player. The KJ game tree has 199 nodes, and 57 sequences per player.

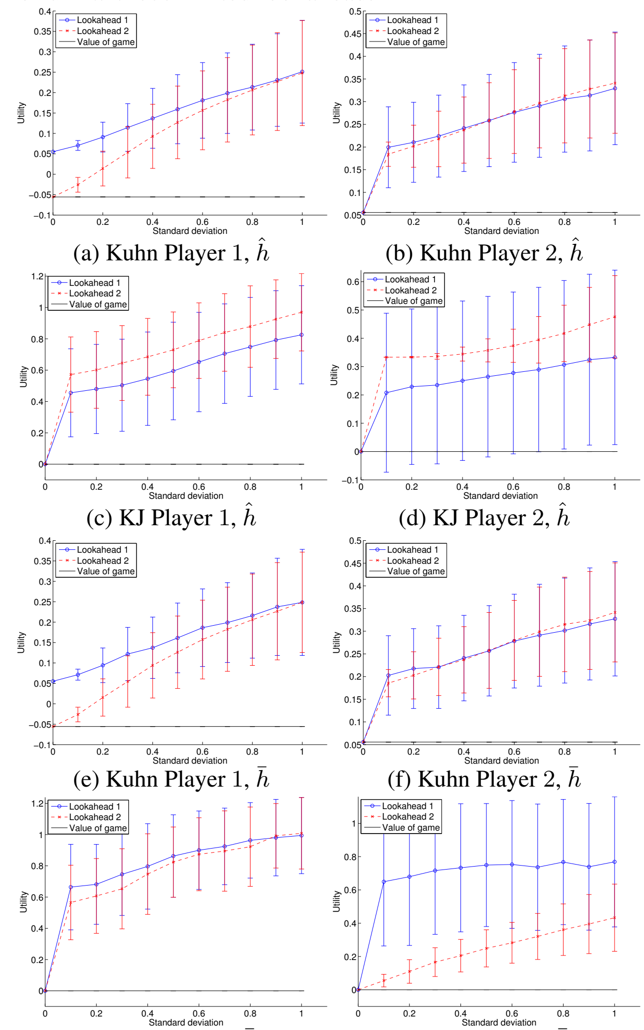
To investigate the value that can be derived from exploiting a limited-lookahead opponent, a node evaluation heuristic is needed. In this work we consider heuristics derived from a Nash equilibrium. For a given node, the heuristic value of the node is simply the expected value of the node in (some chosen) equilibrium. This is arguably a conservative class of heuristics, as a limited-lookahead opponent would not be expected to know the value of the nodes in equilibrium. Even with this form of evaluation heuristic it is possible to exploit the limited-lookahead player, as we will show. We will also consider Gaussian noise being added to the node evaluation heuristic, more realistically modeling opponents who have vague ideas of the values of nodes in the game. Formally, let σ be an equilibrium, and i the limited-lookahead player. The heuristic value $h(s)$ of a node s is:

$$h(s) = \begin{cases} u_i(s) & \text{if } s \in Z \\ \sum_{a \in A_s} \sigma(s,a) h(t^s_a) & \text{otherwise} \end{cases} \tag{15}$$

We consider two different noise models. The first adds Gaussian noise with mean 0 and standard deviation γ independently to each node evaluation, including leaf nodes. Letting μ_s be a noise term drawn from $\mathcal{N}(0, \gamma)$: $\hat{h}(s) = h(s) + \mu_s$. The second, more realistic, model adds error cumulatively, with no error on leaf nodes:

$$\bar{h}(s) = \begin{cases} u_i(s) & \text{if } s \in Z \\ \left[\sum_{a \in A_s} \sigma(s,a) \bar{h}(t^s_a)\right] + \mu_s & \text{otherwise} \end{cases} \tag{16}$$

Using MIP (12), we computed optimal strategies for the rational player in Kuhn poker and KJ. The results are given in Figure 1. The x-axis is the noise parameter γ for $\hat{h}$ and $\bar{h}$. The y-axis is the corresponding utility for the rational player, averaged over at least 1000 runs per tuple $\langle$game, choice of rational player, lookahead, standard deviation$\rangle$. Each figure contains plots for the limited-lookahead player having lookahead 1 or 2, and a baseline for the value of the game in equilibrium without limited lookahead.



(a) Kuhn Player 1, $\hat{h}$ (b) Kuhn Player 2, $\hat{h}$

(c) KJ Player 1, $\hat{h}$ (d) KJ Player 2, $\hat{h}$

(e) Kuhn Player 1, $\bar{h}$ (f) Kuhn Player 2, $\bar{h}$

(g) KJ Player 1, $\bar{h}$ (h) KJ Player 2, $\bar{h}$

Figure 1: Winnings in Kuhn poker and KJ for the rational player as Player 1 and 2, respectively, for varying evaluation function noise. Error bars show standard deviation.

Figures 1a and b show the results for using evaluation function $\hat{h}$ in Kuhn poker, with the rational player in plot a and b being Player 1 and 2, respectively. For rational Player 1, we see that, even with no noise in the heuristic (i.e., the limited-lookahead player knows the value of each node in equilibrium), it is possible to exploit the limited-lookahead player if she has lookahead 1. (With lookahead 2 she achieves the value of the game.) For both amounts of lookahead, the exploitation potential steadily increases as noise is added.

Figures 1c and d show the same variants for KJ. Here, lookahead 2 is worse for the limited-lookahead player than lookahead 1. To our knowledge, this is the first known imperfect-information *lookahead pathology*. Such pathologies are well known in perfect-information games [Beal, 1980; Pearl, 1981; Nau, 1983], and understanding them remains an active area of research [Luštrek, Gams, and Bratko, 2006; Nau et al., 2010; Wilson et al., 2012]. This version of the node heuristic does not have increasing *visibility*: node evaluations do not get more accurate toward the end of the game. Our experiments on KJ with $\bar{h}$ in Figures 1g and h do not have this pathology, and $\bar{h}$ does have increasing visibility.

Figure 2 shows a simple subtree (that could be attached to any game tree) where deeper lookahead can make the agent's decision arbitrarily bad, even when the node evaluation function is the exact expected value of a node in equilibrium.

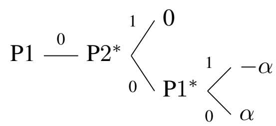


Figure 2: A subtree that exhibits lookahead pathology.

We now go over the example of Figure 2. Assume without loss of generality that all payoffs are positive in some game. We can then insert the subtree in Figure 2 as a subgame at any node belonging to P1, and it will be played with probability 0 in equilibrium, since it has expected value 0. Due to this, all strategies where Player 2 chooses up can be part of an equilibrium. Assuming that P2 is the limited-lookahead player and minimizing, for large enough α, the node labeled P1* will be more desirable than any other node in the game, since it has expected value $-\alpha$ according to the evaluation function. A rational player P1 can use this to get P2 to go down at P2*, and then switch to the action that leads to α. This example is for lookahead 1, but we can generalize the example to work with any finite lookahead depth: the node P1* can be replaced by a subtree where every other leaf has payoff 2α, in which case P2 would be forced to go to the leaf with payoff α once down has been chosen at P2*.

Figures 1e and f show the results for Kuhn poker with $\bar{h}$. These are very similar to the results for $\hat{h}$, with almost identical expected utility for all scenarios. Figures 1g and h, as previously mentioned, show the results with $\bar{h}$ on KJ. Here we see no abstraction pathologies, and for the setting where Player 2 is the rational player we see the most pronounced difference in exploitability based on lookahead.

7 Conclusions and future work

This paper initiated the study of limited lookahead in imperfect-information games. We characterized the complexity of finding a Nash equilibrium and optimal strategy to commit to for either player. Figure 3 summarizes those results, including the cases of favorable and static tie-breaking, the discussion of which we deferred to the extended online paper. We then designed a MIP for computing optimal strategies to

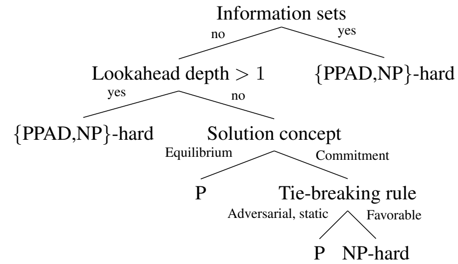


Figure 3: Our complexity results. {PPAD,NP}-hard indicates that finding a Nash equilibrium (optimal strategy to commit to) is PPAD-hard (NP-hard). P indicates polytime.

commit to for the rational player. The problem was shown to reduce to choosing the best among a set of two-player zero-sum games (the tie-breaking being the opponent), where the optimal strategy for any such game can be computed with an LP. We then introduced a MIP that finds the optimal solution by branching on these games.

We experimentally studied the impact of limited lookahead in two poker games. We demonstrated that it is possible to achieve large utility gains by exploiting a limited-lookahead opponent. As one would expect, the limited-lookahead player often obtains the value of the game if her heuristic node evaluation is exact (i.e., it gives the expected values of nodes in the game tree for some equilibrium)—but we provided a counterexample that shows that this is not sufficient in general. Finally, we studied the impact of noise in those estimates, and different lookahead depths. While lookahead 2 usually outperformed lookahead 1, we uncovered an imperfect-information game lookahead pathology: deeper lookahead can hurt the limited-lookahead player. We demonstrated how this can occur with any finite depth of lookahead, even if the limited-lookahead player's node evaluation heuristic returns exact values from an equilibrium.

Our algorithms in the NP-hard adversarial tie-breaking setting scaled to games with hundreds of nodes. For some practical settings more scalability will be needed. There are at least two exciting future directions toward achieving this. One is to design faster algorithms. The other is designing abstraction techniques for the limited-lookahead setting. In extensive-form game solving with rational players, abstraction plays an important role in large-scale game solving [Sandholm, 2010]. Theoretical solution quality guarantees have recently been achieved [Lanctot et al., 2012; Kroer and Sandholm, 2014a; Kroer and Sandholm, 2014b]. Limited-lookahead games have much stronger structure, especially locally around an information set, and it may be possible to utilize that to develop abstraction techniques with significantly stronger solution quality bounds. Also, leading practical game abstraction algorithms (e.g., [Ganzfried and Sandholm, 2014]), while theoretically unbounded, could immediately be used to investigate exploitation potential in larger games. Finally, uncertainty over h is an important future research direction. This would lead to more robust solution concepts, thereby alleviating the pitfalls involved with using an imperfect estimate.

Acknowledgements. This work is supported by the National Science Foundation under grant IIS-1320620.

References

[1] Donald F. Beal. "An analysis of minimax". In: *Advances in computer chess* 2 (1980), pp. 103–109.

[2] Bruno Bouzy and Tristan Cazenave. "Computer Go: an AI oriented survey". In: *Artificial Intelligence* 132.1 (2001), pp. 39–103.

[3] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. "Settling the Complexity of Computing Two-Player Nash Equilibria". In: *Journal of the ACM* (2009).

[4] Vincent Conitzer and Tuomas Sandholm. "Computing the Optimal Strategy to Commit to". In: *Proceedings of the ACM Conference on Electronic Commerce (ACM-EC)*. Ann Arbor, MI, 2006.

[5] Sam Ganzfried and Tuomas Sandholm. "Potential-Aware Imperfect-Recall Abstraction with Earth Mover's Distance in Imperfect-Information Games". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. 2014.

[6] Michael Johanson, Kevin Waugh, Michael Bowling, and Martin Zinkevich. "Accelerating Best Response Calculation in Large Extensive Games". In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 2011.

[7] Richard E. Korf. "Real-time heuristic search". In: *Artificial intelligence* 42.2 (1990), pp. 189–211.

[8] Christian Kroer and Tuomas Sandholm. "Extensive-Form Game Abstraction With Bounds". In: *Proceedings of the ACM Conference on Economics and Computation (EC)*. 2014.

[9] Christian Kroer and Tuomas Sandholm. "Extensive-Form Game Imperfect-Recall Abstractions with Bounds". In: *arXiv preprint: http://arxiv.org/abs/1409.3302* (2014).

[10] Harold W. Kuhn. "A Simplified Two-Person Poker". In: *Contributions to the Theory of Games*. Ed. by H. W. Kuhn and A. W. Tucker. Vol. 1. Annals of Mathematics Studies, 24. Princeton, New Jersey: Princeton University Press, 1950, pp. 97–103.

[11] Marc Lanctot, Richard Gibson, Neil Burch, Martin Zinkevich, and Michael Bowling. "No-Regret Learning in Extensive-Form Games with Imperfect Recall". In: *International Conference on Machine Learning (ICML)*. 2012.

[12] Joshua Letchford and Vincent Conitzer. "Computing Optimal Strategies to Commit to in Extensive-Form Games". In: *Proceedings of the ACM Conference on Electronic Commerce (EC)*. 2010.

[13] Mitja Luštrek, Matjaž Gams, and Ivan Bratko. "Is real-valued minimax pathological?" In: *Artificial Intelligence* 170.6 (2006), pp. 620–642.

[14] Vahab Mirrokni, Nithum Thain, and Adrian Vetta. "A theoretical examination of practical game playing: lookahead search". In: *Algorithmic Game Theory*. Springer, 2012, pp. 251–262.

[15] Dana S. Nau. "Pathology on game trees revisited, and an alternative to minimaxing". In: *Artificial intelligence* 21.1 (1983), pp. 221–244.

[16] Dana S. Nau, Mitja Luštrek, Austin Parker, Ivan Bratko, and Matjaž Gams. "When is it better not to look ahead?" In: *Artificial Intelligence* 174.16 (2010), pp. 1323–1338.

[17] Christos H. Papadimitriou. "On the complexity of the parity argument and other inefficient proofs of existence". In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532.

[18] Judea Pearl. "Heuristic Search Theory: Survey of Recent Results." In: *IJCAI*. Vol. 1. 1981, pp. 554–562.

[19] Judea Pearl. "On the nature of pathology in game searching". In: *Artificial Intelligence* 20.4 (1983), pp. 427–453.

[20] Raghuram Ramanujan, Ashish Sabharwal, and Bart Selman. "On Adversarial Search Spaces and Sampling-Based Planning." In: *ICAPS*. Vol. 10. 2010, pp. 242–245.

[21] Raghuram Ramanujan and Bart Selman. "Trade-Offs in Sampling-Based Adversarial Planning." In: *ICAPS*. 2011, pp. 202–209.

[22] Tuomas Sandholm. "The State of Solving Large Incomplete-Information Games, and Application to Poker". In: *AI Magazine* (2010). Special issue on Algorithmic Game Theory, pp. 13–32.

[23] Tuomas Sandholm. "Steering Evolution Strategically: Computational Game Theory and Opponent Exploitation for Treatment Planning, Drug Design, and Synthetic Biology". In: *AAAI Conference on Artificial Intelligence, Senior Member Track*. 2015.

[24] Bernhard von Stengel. "Efficient Computation of Behavior Strategies". In: *Games and Economic Behavior* 14.2 (1996), pp. 220–246.

[25] Brandon Wilson, Inon Zuckerman, Austin Parker, and Dana S. Nau. "Improving Local Decisions in Adversarial Search." In: *ECAI*. 2012, pp. 840–845.

[26] Z Yin, A Jiang, M Tambe, C Kietkintveld, K Leyton-Brown, T Sandholm, and J Sullivan. "TRUSTS: Scheduling Randomized Patrols for Fare Inspection in Transit Systems using Game Theory". In: *AI Magazine* (2012).