Knowledge-Based Strategies for Multi-Player Games with Imperfect Information

Dilian Gurov (KTH Stockholm), Valentin Goranko (Stockholm University), Edvin Lundberg (Rocker)

ZTSRiO Seminar

10 June 2021

1. The General Problem

How to coordinate a team of intelligent agents to achieve a given joint objective?

Applicable to a **variety of situations** such as:

- A team of robots has to find and remove a dangerous object in a field or building. It can also be a search-and-rescue mission, and the object may be a moving one.
- In an automated factory, a team of robots has to assemble a machine, or in a dangerous field, a housing facility.
- A team of drones has to diagnose a problem in a wind turbine.
- A collection of threads communicating via shared memory has to achieve a common task, e.g., detect intrusion.

The Aspect of "Knowledge"

When the agents have incomplete or **imperfect information** about the state-of-affairs, to achieve a given objective, it is often necessary that they remember (and aggregate) what they observe as they move along.

This leads us to the notion of **knowledge** about the state-of-affairs, as some form of abstraction over observation histories (as in Automata Theory, where we say that states capture *the relevant part of the history*).

But... what should be remembered? I.e., what should this abstraction be, and how shall it be represented, used and updated?

In the presence of multiple agents, we can also consider **higher-order knowledge**. How does it affect the strategic abilities of a team to achieve objectives?

Knowledge-Based Strategies

More generally, the **knowledge** of an agent refers to *information structured suitably for deciding on a course of action towards achieving an objective*. This knowledge can be *static*, e.g., about the arena, or *dynamic*, e.g., about the actions and observations. We shall focus on the latter.

A **knowledge-based strategy** consists of:

1. a **knowledge representation** (for the dynamic knowledge) in a suitable data structure (**knowledge state**, or "state-of-mind");
2. an **action mapping** from knowledge states to actions of the agent;
3. a **knowledge update** function that computes the new knowledge state of the agent from the old one, the action taken, and the observation made upon it.

Closely related to **knowledge-based programs** [Fagin et al. 1997].

2. The Concrete Object of Study

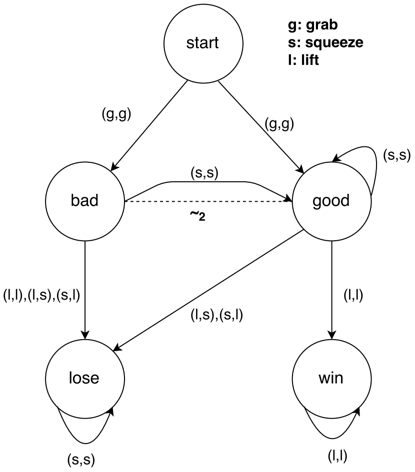
We study the above questions in the context of **multi-player games with imperfect information**. Our investigation starts with the following simplifying **assumptions**:

- the game graph is discrete, finite, and known to the agents,
- all agents are in one team; everything else is modelled as "Nature",
- certain states-of-affairs are indistinguishable for certain agents,
- the agents do not observe each others actions,
- the agents can not communicate with each other,
- the agents should not know each others strategies.

We thus assume that a **central supervisor** will synthesise the individual strategies (i.e., action mappings) of the agents and will distribute them individually before the start of the mission.

Example: Cup-Lifting Robots [Lundberg 2017]

Consider a scenario where two robots must cooperate to lift a cup of acid. The robots must first have a secure overall grip, which can be attained by squeezing, and then both robots must lift at the same time. If not, the cup spills and the game is lost. The first robot has a sensor that indicates whether the grip is good or not, while the second robot doesn't have such a sensor.



Formal Definition of MPGIIAN

Assume a **team** of n players, from player 1 to player n, striving to achieve a common goal.

Definition (Multi-Player Game with Imperfect Information Against Nature)

A **multi-player game graph with imperfect information** is a tuple $G = (L, l_I, \Sigma, \Delta, \mathcal{O})$, where:

(i) L is a finite set of **locations**;

(ii) $l_I \in L$ is the **initial location**;

(iii) $\Sigma = \Sigma_1 \times \ldots \times \Sigma_n$ are the **action profiles** of the team;

(iv) $\Delta \subseteq L \times \Sigma \times L$ are the **transitions**;

(v) $\mathcal{O} = \mathcal{O}_1 \times \ldots \times \mathcal{O}_n$ are the **observation profiles** of the team.

Some Standard Notions in Games over Graphs

A **play** in G is an infinite path $\pi = l_0 l_1 l_2 \ldots$ in G.

Whether a play π in G is **won** by the team is determined by the **objective** of the game. Abstractly, an objective is a set of plays, and a play is won by the team if it is in this set.

An **observable reachability objective** is defined by a set R of individual observations, as the set of plays that visit some element of R.

A **profile of observation-based perfect-recall strategies** for the team is a set $\{\alpha_i\}_{i \leq n}$ of individual strategies $\alpha_i : \mathcal{O}_i^+ \to \Sigma_i$.

The profile is **winning** for a given objective if every play resulting from the players following their individual strategies is in the objective.

Knowledge-Based Subset Construction

For **single-player games with imperfect information**, there is a **Knowledge-Based Subset Construction** (KBSC) [Reif 1984], which translates games of imperfect information to games of perfect information.

The construction is akin to the Subset Construction for converting a NFA to an equivalent DFA, but takes into account actions *and* observations.

Definition (Knowledge-Based Subset Construction)

Let $G = (L, l_I, \Sigma, \Delta, \mathcal{O})$ be a single-player game graph of imperfect information. It induces an (expanded) game graph of perfect information $G^{\mathsf{K}} = (S, s_I, \Sigma, \Delta^{\mathsf{K}})$, where:

(i) $S \stackrel{\text{def}}{=} \left\{ s \in 2^L \setminus \{\varnothing\} \;\middle|\; \exists o \in \mathcal{O}.\ s \subseteq o \right\}$ are the **knowledge states**;

(ii) $s_I \stackrel{\text{def}}{=} \{l_I\}$ is the **initial knowledge state**;

(iii) $\Delta^{\mathsf{K}} \stackrel{\text{def}}{=} \{(s, \sigma, s') \mid \exists o \in \mathcal{O}.\ s' = \{l' \in o \mid \exists l \in s.\ (l, \sigma, l') \in \Delta\}\}$.

Properties of the KBSC

The KBSC translates each single-player game with imperfect information (with a given observable objective) to an "expanded" game of perfect information (with objective), so that a winning observation-based perfect-recall strategy exists in the former game iff a winning perfect-recall one exists in the latter one – where memoryless strategies suffice!

The construction suggests a notion of **knowledge** as a set of locations: a subset of the current observation, representing *the most precise estimate about the current state-of-affairs that the player can deduce*.

This notion of knowledge is thus **sufficient** for strategy synthesis for single-player games with imperfect information, for parity objectives!

From Single-Player to Multi-Player Games

For **single-player games with imperfect information**, and the class of observable parity objectives, an observation-based perfect-recall winning strategy exists if and only if a finite-memory one exists [Chatterjee et al. 2007].

For **multi-player games with imperfect information**, the problem of whether a team has an observation-based perfect-recall winning strategy is **undecidable** even just for the class of observable reachability objectives [Pnueli & Rosner 1990].

This implies that there *cannot* be a strategy-preserving translation to (finite) games with perfect information. (There is a so-called Epistemic Unfolding by Berwanger et al. that potentially yields an infinite game.)

3. Higher-Order Knowledge & Knowledge-Based Strategies

We define **higher-order knowledge** of a player inductively:

- **Order-0 knowledge** is defined by what the player currently senses, i.e., the current state-of-affairs as registered by the sensors. Represented as an observation, i.e., as a set of locations.
- **Order-1 knowledge** is *the most precise estimate about the current state-of-affairs that the player can deduce*. Represented as a non-empty subset of an observation, i.e., as a set of locations.
- **Order-(k+1) knowledge** consists of the order-1 knowledge of the player and the possible order-k knowledge of all other players.

An **order-k knowledge-based strategy** of a player maps individual order-k knowledge states of the player to actions of that player. Existence of a winning one is decidable for a fixed k.

Knowledge Update

As individual **order-1 knowledge update** functions we define:

$$\delta_i(s, \sigma_i, o_i) \stackrel{\text{def}}{=} \left\{ l' \in o_i \;\middle|\; \exists \sigma \in \Sigma.\ (\sigma(i) = \sigma_i \wedge \exists l \in s.\ (l, \sigma, l') \in \Delta) \right\}$$

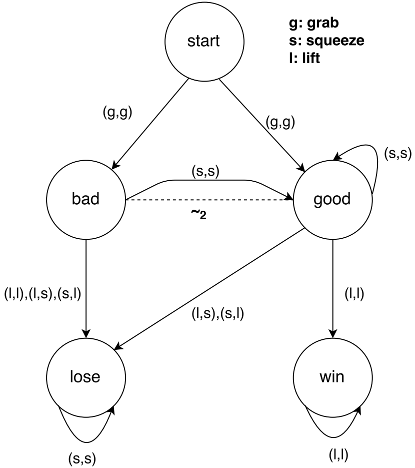
where $s \subseteq L$, $\sigma_i \in \Sigma_i$ and $o_i \in \mathcal{O}_i$.

As individual **order-(k+1) knowledge update** functions we define:

$$\delta_i^{k+1}(s, \sigma_i, o_i) \stackrel{\text{def}}{=} \left\{ (\delta_1^k(A_1, \sigma_1, o_1), \ldots, \delta_n^k(A_n, \sigma_n, o_n)) \;\middle|\; \begin{array}{l} (A_1, \ldots, A_n) \in s, \\ (\sigma_1, \ldots, \sigma_n) \in \Sigma, \\ (o_1, \ldots, o_n) \in \mathcal{O}^p \end{array} \right\}$$

where $s \in A^{(k+1)}$, $\sigma_i \in \Sigma_i$ and $o_i \in \mathcal{O}_i$.

Order-0 Knowledge-Based Winning Strategy?



Robot 1 needs to decide for the observations:

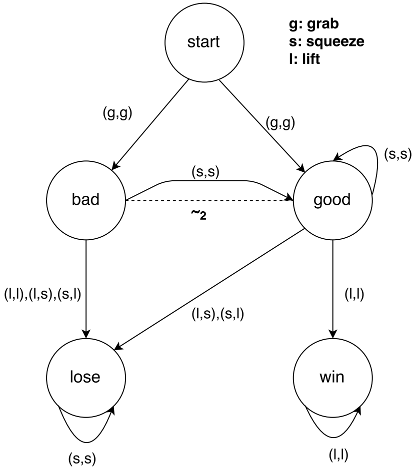
$$\{\text{bad}\} \longrightarrow \text{squeeze}$$
$$\{\text{good}\} \longrightarrow \text{lift}$$

while Robot 2 needs to decide for the observation:

$$\{\text{bad}, \text{good}\} \longrightarrow \text{do what?}$$

No choice guarantees a win.

Order-1 Knowledge-Based Winning Strategy?



Robot 1 needs to decide for the knowledge states:

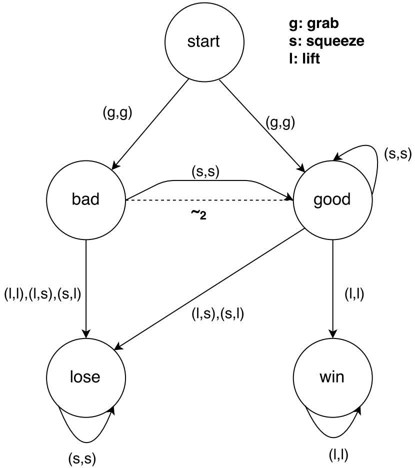
$$\begin{array}{rcl} \{\text{bad}\} & \longrightarrow & \text{squeeze} \\ \{\text{good}\} & \longrightarrow & \text{lift} \end{array}$$

while Robot 2 needs to decide for the knowledge states:

$$\begin{array}{rcl} \{\text{bad}, \text{good}\} & \longrightarrow & \text{do what?} \\ \{\text{good}\} & \longrightarrow & \text{lift} \end{array}$$

No choice guarantees a win.

Order-2 Knowledge-Based Winning Strategy



Robot 1 picks:

$$\{(\{\text{bad}\}, \{\text{bad}, \text{good}\})\} \longrightarrow \text{squeeze}$$

$$\{(\{\text{good}\}, \{\text{bad}, \text{good}\})\} \longrightarrow \text{squeeze}$$

$$\{(\{\text{good}\}, \{\text{good}\})\} \longrightarrow \text{lift}$$

while Robot 2 picks:

$$\{(\{\text{bad}\}, \{\text{bad}, \text{good}\}), (\{\text{good}\}, \{\text{bad}, \text{good}\})\} \longrightarrow \text{squeeze}$$

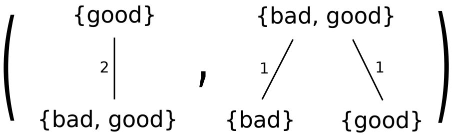
$$\{(\{\text{good}\}, \{\text{good}\})\} \longrightarrow \text{lift}$$

Interpreting Order-2 Knowledge

The order-2 **joint knowledge state**:

$$(\{(\{\text{good}\}, \{\text{bad}, \text{good}\})\}, \ \{(\{\text{bad}\}, \{\text{bad}, \text{good}\}), \ (\{\text{good}\}, \{\text{bad}, \text{good}\})\})$$

can be represented as a pair of trees:



Closely related to the *k*-trees of [van der Meyden 1998].

4. Synthesis of Knowledge-Based Strategies

Our approach:

1. From the given game, compute another, **expanded game**, with joint knowledge states as locations. Translate the given objective.
2. In the expanded game, search for **memoryless observation-based strategy**.
3. If found, present the strategy as order-1 knowledge-based strategy for play in the original game.

We generalise the KBSC, which computes the order-1 knowledge states of a player, to the multi-agent setting (MKBSC). As already explained, the resulting game is in general also a game with imperfect information.

Formal Definition of the MKBSC

Definition (MKBSC)

1. **Projection**: $G|_i \stackrel{\text{def}}{=} (L, l_I, \Sigma_i, \Delta_i, \mathcal{O}_i)$, where:
   $(l, \sigma_i, l') \in \Delta_i \stackrel{\text{def}}{\iff} \exists \sigma \in \Sigma.\ (\sigma(i) = \sigma_i \wedge (l, \sigma, l') \in \Delta)$.
2. **Expansion**: the standard KBSC.
3. **Composition**: synchronous product, followed by **pruning**: remove inconsistent knowledge states (empty intersection) and unrealisable transitions (existential abstraction).
4. **Partition**: $s \sim_i s' \stackrel{\text{def}}{\iff} s(i) = s'(i)$.

First proposed in [Lundberg 2017], later presented in current form.

Implementation available from `github.com/helmernylen/mkbsc`, described in [Nylén & Jacobsson 2018].

Stabilisation and Recursive Knowledge

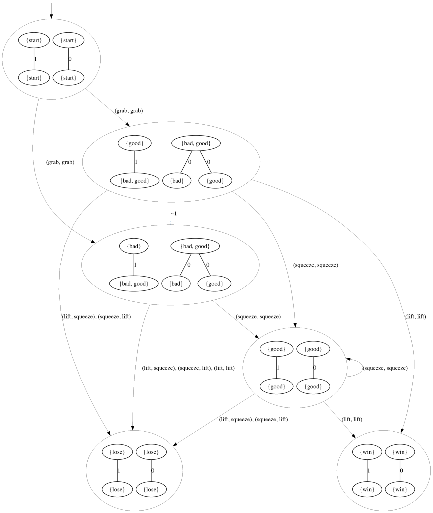
The MKBSC has to yield a game of imperfect information, and can thus be *iterated*! Iterating the MKBSC k-times computes order-k knowledge states.

Interestingly, on some games the iterated construction **stabilises**, in the sense that it results in isomorphic games! For instance, for the above game graph, G^{3K} is isomorphic to G^{2K}.

This leads to a number of interesting questions about stabilisation and the properties of stable game graphs. Are there structural conditions that are sufficient for a game to be "eventually stable"?

In particular, the knowledge encoded in the locations of stable game graphs can be "folded" into recursive representations. This allows the representation of **common knowledge**.

The Order-2 Expansion G^{2K} of the Cup-Lifting Game G



A Strategy Synthesis Procedure

To synthesise knowledge-based strategies:

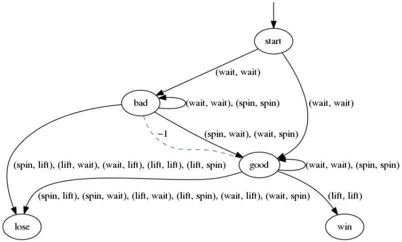
repeat
  apply MKBSC on latest game;
  search for observation-based memoryless strategy in expanded game;
until
  strategy found or
  game repeats.

If the procedure terminates at iteration k with a strategy found, convert the latter to order-k knowledge-based strategy.

If the procedure terminated because of stabilisation, no knowledge-based strategy exists of any order.

Limitations of the MKBSC

The following game graph is stable:



There is no observation-based memoryless winning strategy, and hence no winning knowledge-based strategy of any order. But there is a finite-memory winning strategy.

The class of knowledge-based strategies is thus strictly weaker than the class of finite-memory ones.

5. Conclusions

Higher-order **knowledge** can be based on the notion of **most precise estimate of the current state-of-affairs**. (But other notions are also possible.)

The search for order-k knowledge-based strategies can be reduced to the search for **observation-based memoryless strategies** in expanded games.

In general, the higher the **knowledge order**, the higher the **strategic abilities** of the team. But... we may encounter the phenomenon of **saturation** of knowledge, which manifests itself as stabilisation of the MKBSC.

If the iterated MKBSC stabilises and there is no winning memoryless observation-based strategy in the stable expansion, then there is no knowledge-based strategy of any order in the original game.

Ongoing and Future Work

Characterise the class of objectives that can be achieved with knowledge-based strategies as defined here (**strategic ability**).

Exploit the definition of the MKBSC to design efficient **strategy synthesis** algorithms.

Find sufficient conditions for **stabilisation** and **non-stabilisation** of the iterated MKBSC.

Relax selected **assumptions**. Define corresponding expansions. Compare strategic abilities.

Explore other **knowledge representations**. Define corresponding expansions.

Explore **temporal logic** for defining objectives and **epistemic logic** for presenting knowledge-based strategies.

Evaluate the **practical utility** of the method.

Available Material

This work is currently under review. A **report** is available on arXiv:

> *Dilian Gurov, Valentin Goranko, Edvin Lundberg:*
> *Knowledge-Based Strategies for Multi-Agent Teams Playing Against Nature.*
> *CoRR abs/2012.14851 (2021)*

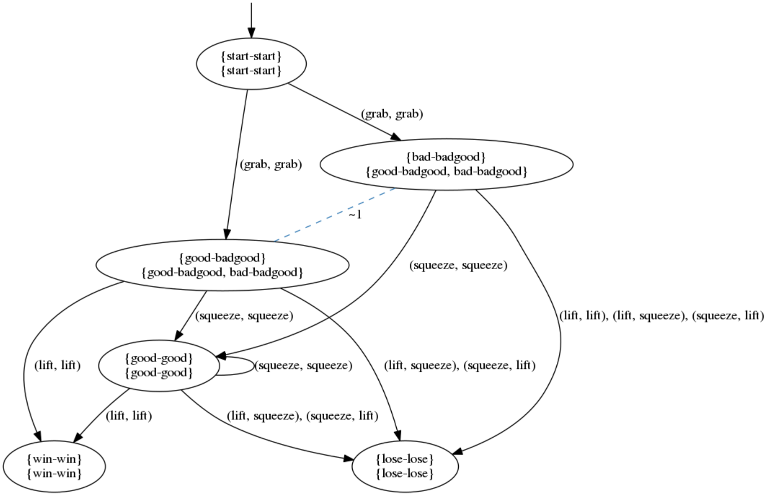
and a **tool** on GitHub implementing the MKBSC:

`github.com/helmernylen/mkbsc`

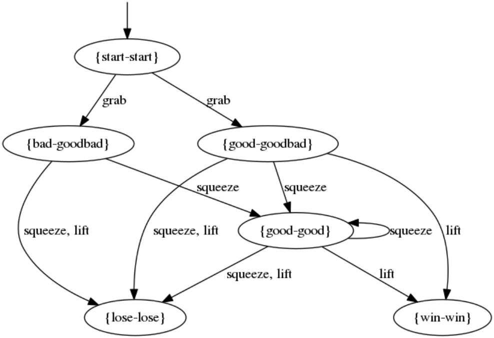
Appendix

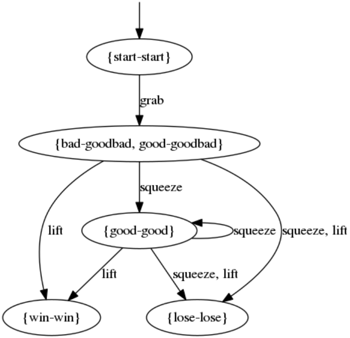
Supplementary Material

The Order-2 Expansion of the Cup-Lifting Game



The Individual Views





The previous expansion is the product of these two individual expansions, which are **single-player games with perfect information**. Furthermore, individual observation-based strategies in the former game are strategies in the latter games. This makes them useful for **strategy synthesis**.