Lifting in Multi-agent Systems under Uncertainty

Tanya Braun[1], Marcel Gehrke[2], Florian Lau[2], Ralf Möller[2]
[1]Computer Science Department, University of Münster,
[2]Institute of Information Systems, University of Lübeck, [3] Institute of Telematics, University of Lübeck

Multi-agent Systems

- Joint state and reward for transition, sensor, reward functions
- Individual actions and observations

Complexity exponential in number of agents n

$\Downarrow$

Problems with large n not computable, e.g., nano-scale systems

Under Symmetries

- Agent types: same actions and observations available
- Additional symmetries in functions

Best case symmetry: worst-case dependency down to $\log n$ only

$\Downarrow$

Reuse existing algorithms for types; new task: optimise w.r.t. n

Lifting in Multi-agent Systems under Uncertainty

Tanya Braun[1], Marcel Gehrke[2], Florian Lau[2], Ralf Möller[2]
[1]University of Münster, Münster, Germany
[2]University of Lübeck, Lübeck, Germany

Link to preliminary paper (QR code): https://www.ifis.uni-luebeck.de/uploads/tx_wapublications/braun_143_public.pdf
Please cite full version when published by PMLR.

DecPOMDP

- *Decentralised Partially Observable Markov Decision Problem*
  - *Set of agents* working towards a *joint reward*
  - Environment a *Markov-1 stochastic process*
- *DecPOMDP model M*

$$\left(\boldsymbol{I}, S, \{A_i\}_{i=1}^N, T, R, \{O_i\}_{i=1}^N, \Omega\right)$$

  - $\boldsymbol{I}$ a set of agents, $|\boldsymbol{I}| = N$,
  - S a random variable for the state space,
  - A_i a decision variable; $\boldsymbol{A} = \times_{i=1}^N ran(A_i)$ set of joint actions
  - $T(S', S, \boldsymbol{A}) = P(S' \mid S, \boldsymbol{A})$ a transition model,
  - $R(S)$ a reward function,
  - O_i a random variable; $\boldsymbol{O} = \times_{i=1}^N ran(O_i)$ set of joint observations,
  - $\Omega(\boldsymbol{O}, S) = P(\boldsymbol{O} \mid S)$ a sensor model.
  - Given horizon τ, each agent i has a local policy π_i : $ran(O_{i,(0:t)}) \mapsto ran(A_i); t < \tau; \boldsymbol{\pi} = (\pi_i)_{i=1}^N$ a joint policy
- *DecPOMDP Semantics*: all possible joint policies Π_M
- *DecPOMDP Problem*: find joint policy $\boldsymbol{\pi}^*$ that maximises the expected utility with discount factor $\gamma \in [0,1]$

$$U_M^{\boldsymbol{\pi}}(s_t, \boldsymbol{o}_{0:t}) = R\big(s_t, \underbrace{\boldsymbol{\pi}(\boldsymbol{o}_{0:t})}_{\boldsymbol{a}_{0:t}}\big) + \gamma^t \sum_{s_{t+1} \in ran(S)} T\big(s_{t+1}, s_t, \boldsymbol{\pi}(\boldsymbol{o}_{0:t})\big) \cdot \sum_{\boldsymbol{o}_{t+1} \in ran(\boldsymbol{O})} \Omega(\boldsymbol{o}_{t+1}, s_{t+1}) U_M^{\boldsymbol{\pi}}(s_{t+1}, \boldsymbol{o}_{0:t+1})$$

Problem

- **Complexities** $\mathbb{T}, \mathbb{R}, \mathbb{O}$ of T, R, Ω, evaluation cost $\mathbb{C}$ of a joint policy and size of policy space $\mathbb{P}$ ***exponential*** in N

$$\mathbb{T} \in O(s^2 a^N) \quad \mathbb{R} \in O(s a^N) \quad \mathbb{O} \in O(s o^N)$$
$$\mathbb{C} \in O(s o^N \tau) \quad \mathbb{P} \in O(a^{\frac{N(o^\tau - 1)}{o-1}})$$

  - $s = |ran(S)|, a = \max_i |ran(A_i)|, o = \max_i |ran(O_i)|$

→ Especially a problem with $N \gg 10{,}000$

Symmetric & Partitioned DecPOMDP

- *Symmetric* DecPOMDP: $K \ll N$ *partitions* in agent set
  - In each partition: the same set of actions and observations
  - Formally, $\boldsymbol{I} = \bigcup_{k=1}^K \Im_k, \Im_k \neq \emptyset, \Im_k \cap \Im_{k'} = \emptyset$; for each $\Im_k$: $\forall i, j \in \Im_k : ran(A_i) = ran(A_j) \wedge ran(O_i) = ran(O_j)$
- *Partitioned* model

$$\left(\bar{\boldsymbol{I}}, S, \{\bar{A}_k\}_{k=1}^K, \bar{T}, \bar{R}, \{\bar{O}_k\}_{k=1}^K, \bar{\Omega}\right)$$

  - $\bar{\boldsymbol{I}}$ a partitioning $\{\Im_k\}_{k=1}^K$ of an agent set, $n_k = |\Im_k|$, $|\bar{\boldsymbol{I}}| = N$,
  - $\bar{A}_k, \bar{O}_k$ variables per partition, with joint actions and observations defined over the K partitions
  - $\bar{T}, \bar{R}, \bar{\Omega}$ like T, R, Ω but defined over partitioned inputs

Counting DecPOMDP

- Indistinguishability of agents in a partition yields invariance towards which particular partition agents perform an action or observe an event: it only matters how many agents do or observe something
  - E.g., 2 actions performed by 5 agents each: 10 over 5 and 5 = 255 different permutations of 10 agents to do the actions with the same outcome

→ Count occurrences $[n_1, \dots, n_l], l = |ran(A_k)|$

- *Counting DecPOMDP*: a partitioned DecPOMDP with
  - $\bar{A}_k = \#_{X_k}[A_k(X_k)]$ a counting random variable
  - $\bar{O}_k = \#_{X_k}[O_k(X_k)]$ a counting random variable
  - $\bar{T}, \bar{R}, \bar{\Omega}$ defined over the counting random variables

$$\left(\bar{\boldsymbol{I}}, S, \{\#_{X_k}[A_k(X_k)]\}_{k=1}^K, \bar{T}, \bar{R}, \{\#_{X_k}[O_k(X_k)]\}_{k=1}^K, \bar{\Omega}\right)$$

- **Theorem**:
  Counting model M_c has an equivalent ground model $\mathrm{gr}(M_c)$.
- **Theorem**:
  Optimal policy in M_c also optimal in $\mathrm{gr}(M_c)$.
- **Complexities** $\mathbb{T}_c, \mathbb{R}_c, \mathbb{O}_c, \mathbb{C}_c$ ***polynomial***, $\mathbb{P}_c$ ***exponential*** in $n < N$

$$\mathbb{T}_c \in O(s^2 n^{Ka}) \quad \mathbb{R}_c \in O(s n^{Ka}) \quad \mathbb{O}_c \in O(s n^{Ko})$$
$$\mathbb{C}_c \in O(s n^{K\tau o}) \quad \mathbb{P}_c \in O(n^{a^{\frac{K(n^{\tau o}-1)}{n^o-1}}})$$

  - $n = \max_k n_k, |ran(\bar{A}_k)| \leq n^a, |ran(\bar{O}_k)| \leq n^o$

Isomorphic DecPOMDP

- Independence assumption between agents of a partition
  → $\bar{T}, \bar{R}, \bar{\Omega}$ factorise identically for each agent within each partition
  → Higher efficiency at the expense of lower expressiveness
- *Isomorphic DecPOMDP*: a partitioned DecPOMDP with
  - $\bar{A}_k = A_k(X_k)$ a parameterised random variable
  - $\bar{O}_k = O_k(X_k)$ a parameterised random variable
  - $\bar{T}, \bar{R}, \bar{\Omega}$ defined over the parameterised random variables

$$\left(\bar{\boldsymbol{I}}, S, \{A_k(X_k)\}_{k=1}^K, \bar{T}, \bar{R}, \{O_k(X_k)\}_{k=1}^K, \bar{\Omega}\right)$$

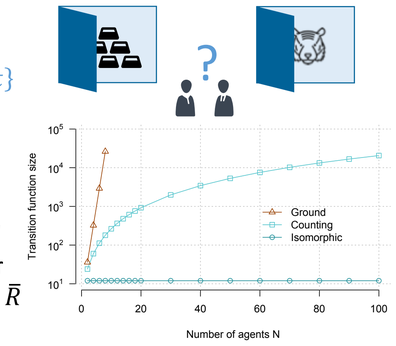
- **Lemma**:
  Isomorphic model M_i has an equivalent counting model M_c.
- **Lemma**:
  Enough to define policies in M_i over $ran(A_k)$ and $ran(O_k)$.
- **Corollary**:
  Partition sizes only influence $U_M^{\boldsymbol{\pi}}$, not $\boldsymbol{\pi}^*$ itself.
- **Complexities** $\mathbb{T}_i, \mathbb{R}_i, \mathbb{O}_i, \mathbb{C}_i, \mathbb{P}_i$ ***logarithmic*** in $n < N$

$$\mathbb{T}_i \in O(s^2 a^K) \quad \mathbb{R}_i \in O(s a^K) \quad \mathbb{O}_i \in O(s o^K)$$
$$\mathbb{C}_i \in O(\log_2(n)\, s o^{K\tau}) \quad \mathbb{P}_i \in O(a^{\frac{K(o^\tau - 1)}{o-1}})$$

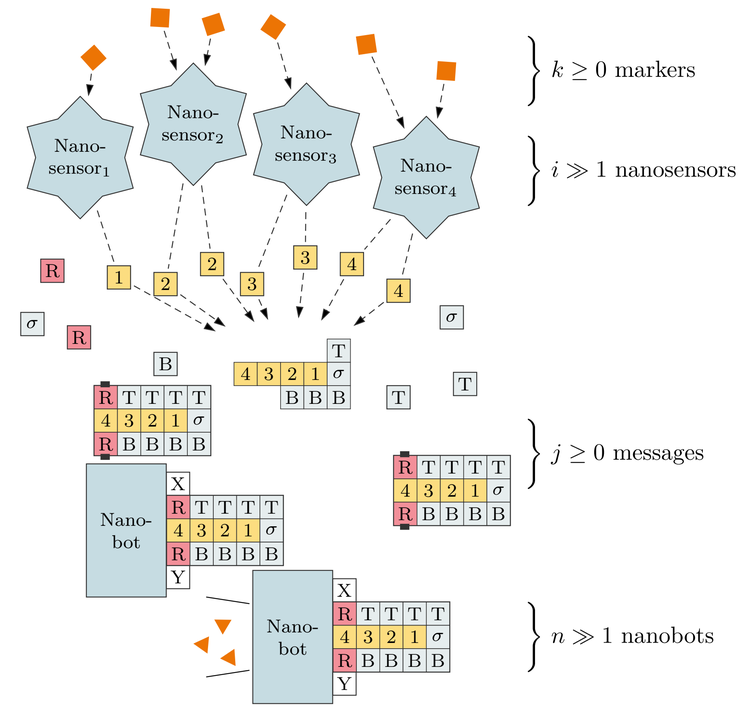
Possible to reuse existing solution approaches to solve DecPOMDP for K partitions; adapt result w.r.t. sizes n_k to reach a given U

Example: DecTiger

- $\boldsymbol{I} = \{agent_1, agent_2\}, \Im_{K=1} = \boldsymbol{I}$
- $ran(S) = \{tigerleft, tigerright\}$
- $A_K(X_K) = \{listen, openleft, openright\}$
- $O_K(X_K) = \{hearleft, hearright\}$
- As a counting model:
  T, R, Ω equivalently encoded in $\bar{T}, \bar{R}, \bar{\Omega}$
- As an isomorphic model:
  - Reset in T not possible to encode in $\bar{T}$
  - Agreeing on an action yielding a lower punishment not possible to encode in $\bar{R}$
  - Ω equivalently encoded in $\bar{\Omega}$



Application: Nanoscale Medical Systems



- Agent set: $\kappa = 4$ different types of markers/sensors, $\iota = 1$ different types of messages/nanobots; $\kappa + \iota = K$, n_k~64,000 (preliminary experiments)
- Each $\Im_k$: 2 actions (output, no act.), 2 observations (sense/receive, no obs.)

WWU MÜNSTER
UNIVERSITÄT ZU LÜBECK