# Novel Class of Expected Value Bounds and Applications in Belief Space Planning

**Ohad Levy-Or**

TECHNION
Israel Institute of Technology

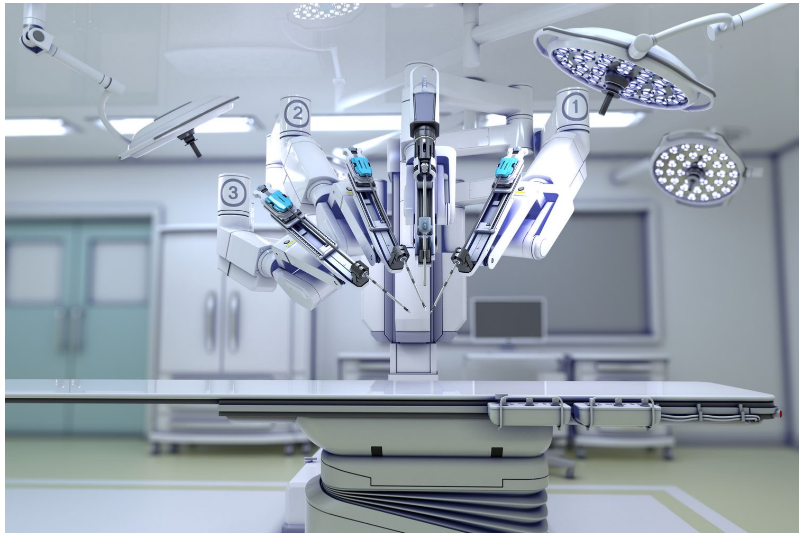
TASP | TECHNION AUTONOMOUS SYSTEMS PROGRAM

ANPL
Autonomous Navigation and Perception Lab

August 6th, 2024

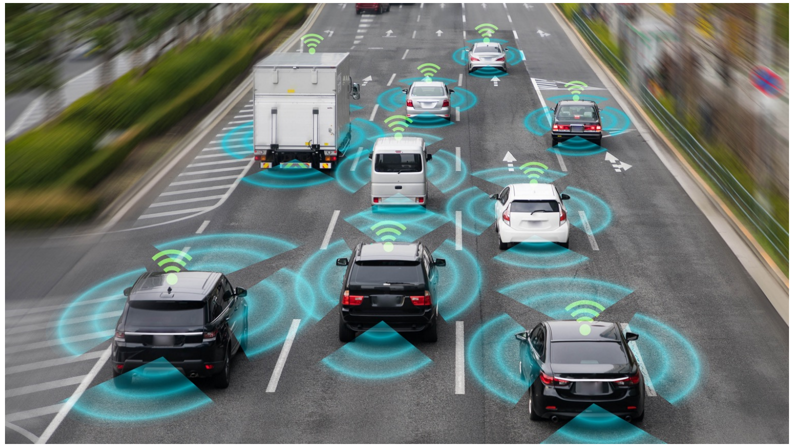
# Introduction
## Autonomy

Drones



Medical Robots

Cars



# Introduction
## Uncertainty

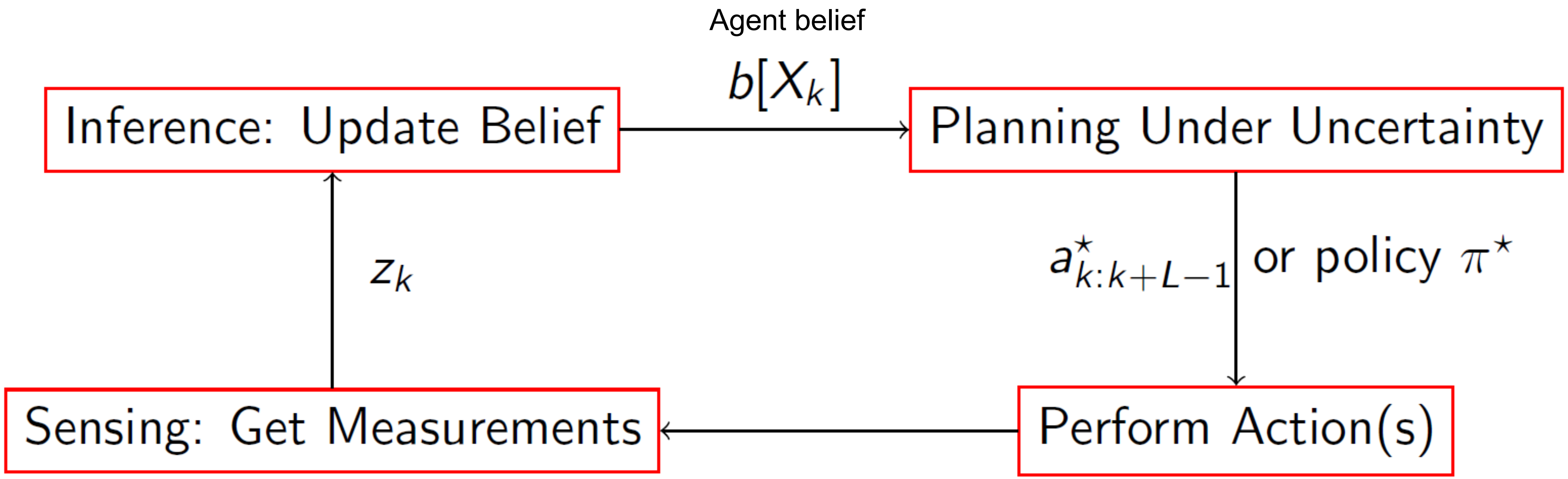
- Imprecise actions
- Imprecise measurements
- Changing environment

An uncertain agent

# Introduction
## Challenges

- Solving problems online
- Computational limitations

# Planning Framework




- Agent **cannot directly observe the state**, so it must **maintain a belief** over its state
- The current agent belief is used as the initial belief for the planner

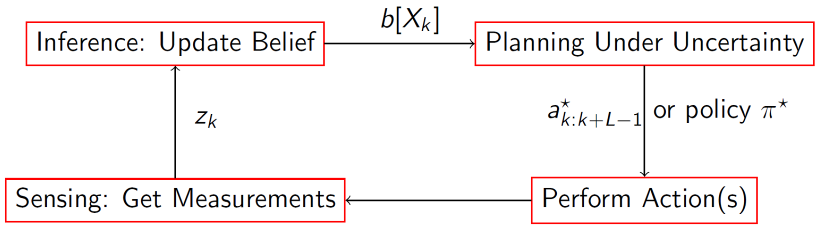
# Probabilistic Inference

- Prediction step:

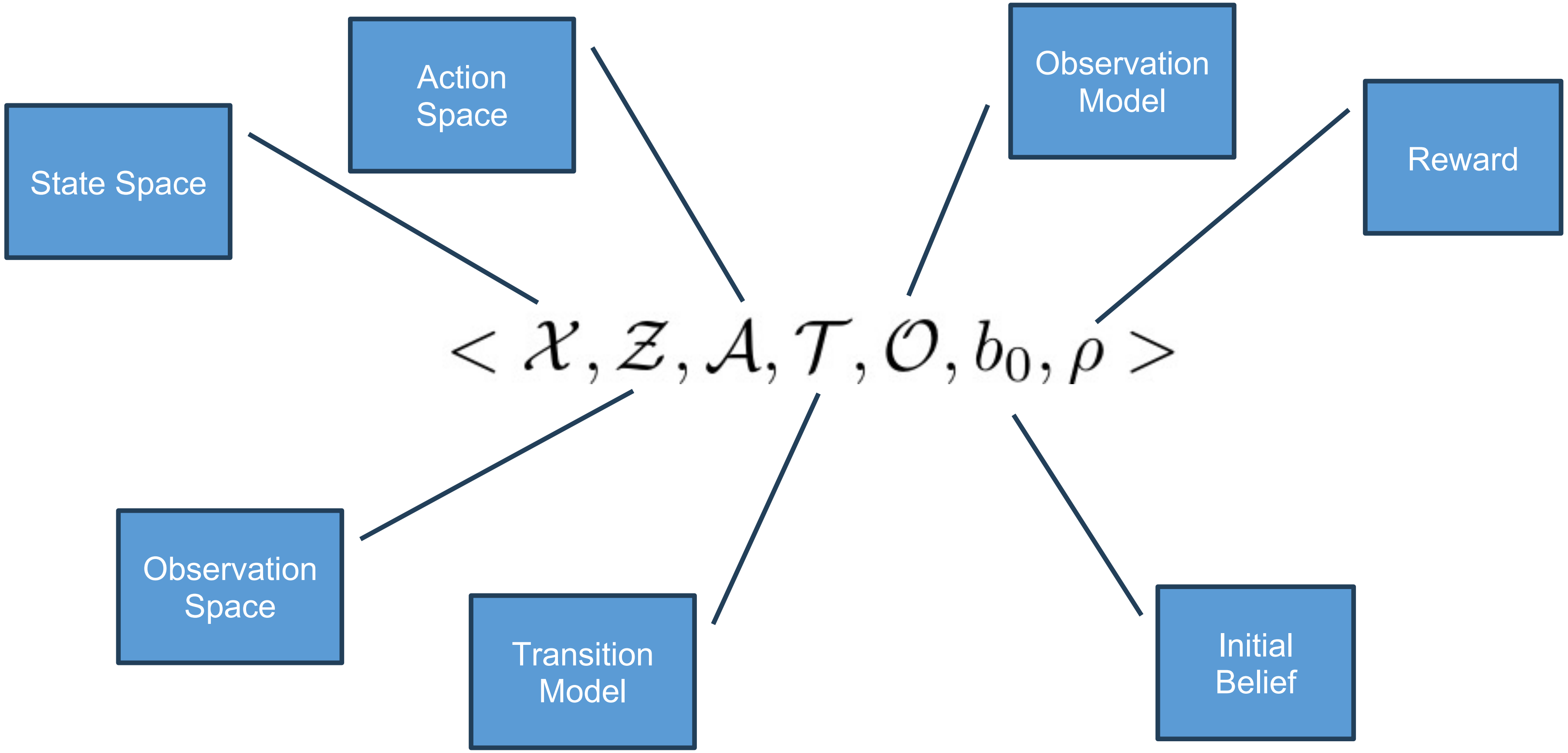
$$b_{k+1}^{-} = \int \mathrm{P}\left(X_{k+1} \mid X_k, a_k\right) b_k \mathrm{d}X_k$$

- Update step:

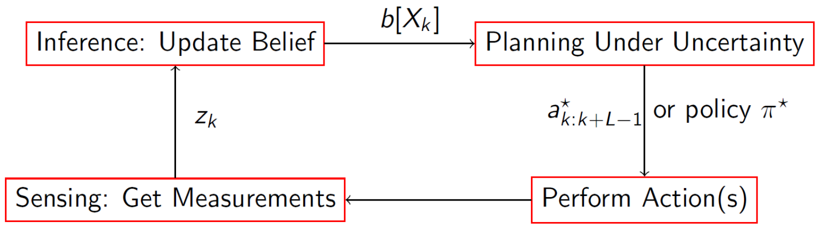
$$b_{k+1} = \eta_{k+1}^{-1} \mathrm{P}\left(Z_{k+1} \mid X_{k+1}\right) b_{k+1}^{-}$$
$$\eta_k \triangleq \mathrm{P}\left(Z_k \mid H_k^{-}\right)$$



# POMDP Formulation



# Belief Space Planning




- We wish to find a policy/action sequence that minimizes the value function:

$$V^{\pi}(b_k) = \underset{\boldsymbol{Z}_{k+1}}{\mathbb{E}} \left[\rho\left(b_k, \pi_k(b_k)\right)\right] + \gamma \underset{\boldsymbol{Z}_{k+1}}{\mathbb{E}} \left[V^{\pi}(b_{k+1})\right]$$

- The general reward $\rho$ may be either state dependent:

$$\underset{\boldsymbol{X}\sim b}{\mathbb{E}} [R(\boldsymbol{X}, \pi(b))]$$

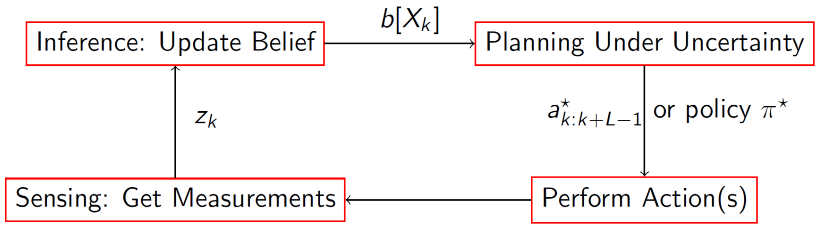
- Or belief dependent:

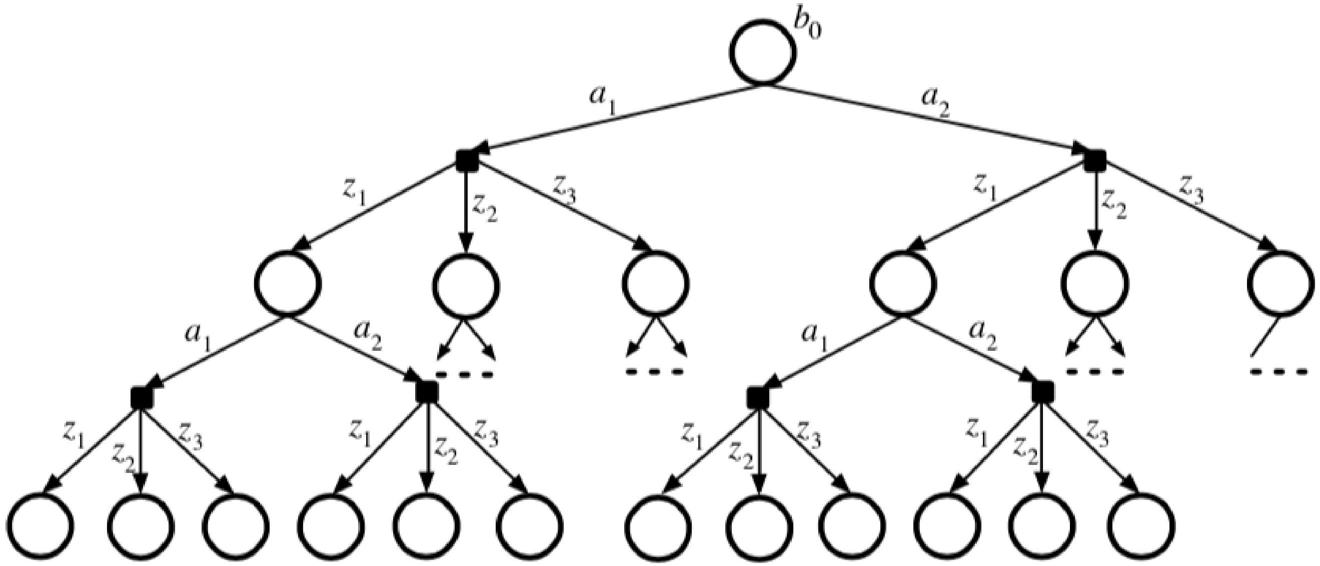
$$\underset{\boldsymbol{X}\sim b}{\mathbb{E}} [R(b(\boldsymbol{X}), \boldsymbol{X}, \pi(b))]$$

All these expectations require some method to calculate them (analytically, Monte Carlo, etc.)

# Belief Tree

- A data structure created to assist in reasoning about the optimal action
- Expands actions and observations via an algorithm of choice up to the defined horizon



# Challenges

- Finding a globally optimal solution is intractable
- Many prior works try to find asymptotically optimal solutions

What can we do?

Simplification with guarantees!

# Simplification with Guarantees

- Find a simpler problem that is tractable to calculate
- Use known bounds to bound the difference between the simplified and unsimplified results
- Uses:
  - Quantify the suboptimality of an algorithm
  - Anytime guarantees on an algorithm (deterministic bounds)
  - If the bounds are tight enough, the optimal actions can still be selected
- When using probabilistic bounds, simplification comes with probabilistic guarantees

# Probability Theory Bounds

- Markov Inequality:

$$P(\boldsymbol{S} \geq a) \leq \frac{E[\boldsymbol{S}]}{a}$$

- Hoeffding’s Inequality:

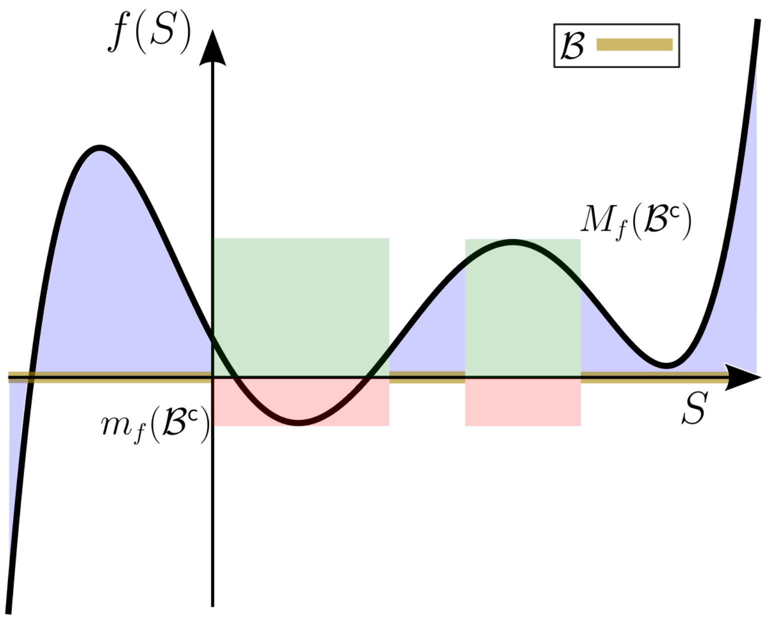
$$P\left(\left|E[f(\boldsymbol{S})] - \frac{\sum_{i=1}^{n} f(\boldsymbol{S}_i)}{n}\right| \leq \sqrt{\frac{\Delta_f^2}{2n}\log\frac{2}{\delta}}\right) \geq 1 - \delta$$

# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer's estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar belief topologies when calculating conditional entropy

# Partial Expectation

- What if we could simplify the problem by only calculating the expectation on part of the support?
- Normal expectation ( $\mathbb{E}\left[f(\boldsymbol{S})\right]$ ) requires integration over the **entire support**
- Partial expectation ( $\mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right]$ ) requires integration over **part of the support**
    - Support is split into $\mathcal{B} \cup \mathcal{B}^{\mathsf{c}} = \mathcal{S}$




$$\mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \triangleq \int_{\mathcal{S}} \mathrm{P}\left(S\right) f(S) \mathbb{1}\{S \in \mathcal{B}\} \mathrm{d}S \equiv \mathbb{E}\left[f(\boldsymbol{S}) \mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\}\right]$$

# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer’s estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Markov Bound Generalization

- Lower bound:

$$\mathbb{E}\left[f(\boldsymbol{S})\right] \geq \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}^{\mathsf{c}}\right) m_f(\mathcal{B}^{\mathsf{c}})$$

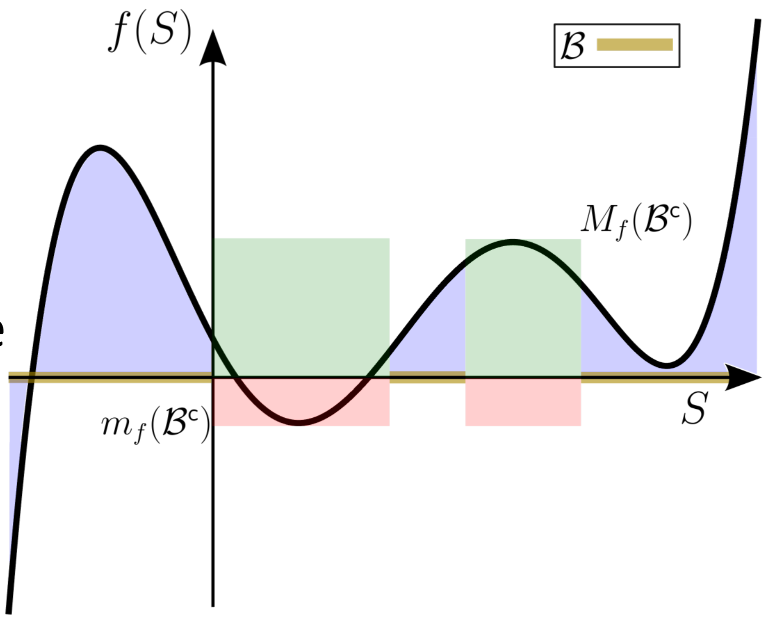
- Upper bound:

$$\mathbb{E}\left[f(\boldsymbol{S})\right] \leq \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}^{\mathsf{c}}\right) M_f(\mathcal{B}^{\mathsf{c}})$$

- $\mathcal{B}^{\mathsf{c}}$ can be further split into multiple subsets for a more general bound

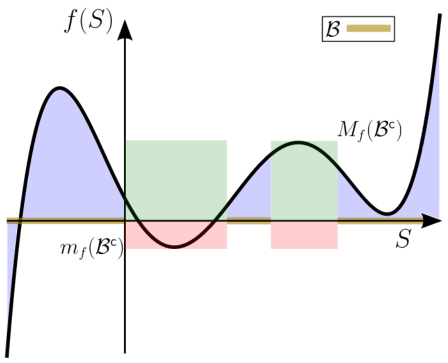
$$\mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}\right) \equiv \mathbb{P}\left(\mathcal{B}\right)$$

Measure of the subset



# Markov Bound Generalization

- Assume $f(S) \geq 0 \rightarrow \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \geq 0$
- This quickly leads to the Markov inequality:



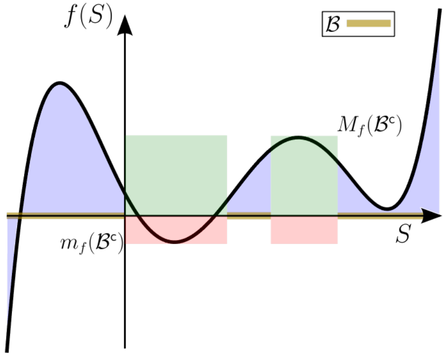

$$\mathbb{E}\left[f(\boldsymbol{S})\right] \geq \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}^c\right) m_f\left(\mathcal{B}^c\right)$$

$\geq 0$

$$\frac{\mathbb{E}\left[f(\boldsymbol{S})\right]}{m_f\left(\mathcal{B}^c\right)} \geq \mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}^c\right)$$

# Markov Bound Generalization

- Properties:
  - **Convergence** to the expected value
  - **Partial incrementality**, reuse loose bounds to get tighter bounds
  - **Monotonicity** of the difference between expectation and partial expectation
- Beneficial for planning algorithms



# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer's estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Hoeffding Bound Generalization

- Blending Hoeffding's inequality with our novel inequality yields two novel probabilistic bounds
- If we sample from $S$ and then select a **subset of samples**:

$$\mathrm{P}\left(\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \hat{\mathcal{E}}_{\mathcal{B}_n}\left[f(\boldsymbol{S})\right] \leq \mathcal{UB}\right) \geq 1-\delta \qquad \forall\delta \in (0,1)$$

- If we **sample from a subspace**:

$$\mathrm{P}\left(\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \mathbb{P}\left(\mathcal{B}\right)\hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] \leq \mathcal{UB}\right) \geq 1-\delta \qquad \forall\delta \in (0,1)$$

- ^ indicates empirical estimator

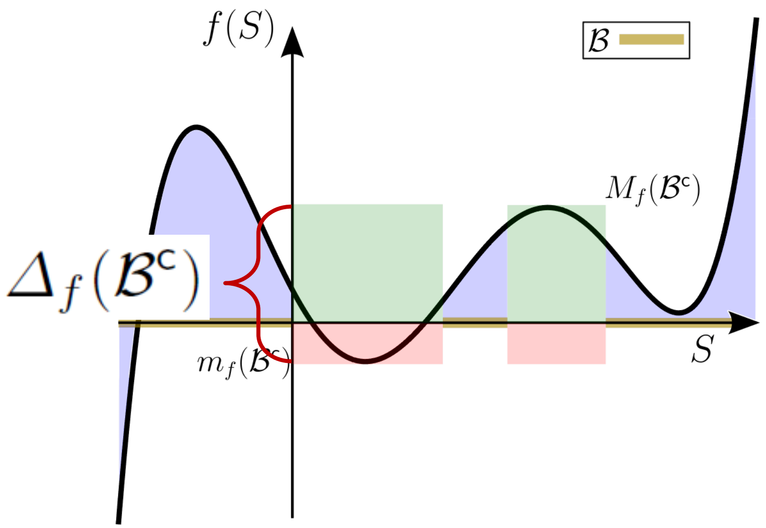
# Hoeffding Bound Generalization

- Conditions exist that guarantee improvement upon Hoeffding's inequality
  - i.e the bound gap (upper bound – lower bound) is smaller
- These conditions depend not only on the measure of $\mathcal{B}$ but on the actual choice.

$$C \cdot \left(\Delta_f - \mathbb{P}(\mathcal{B})\,\Delta_f(\mathcal{B})\right) \geq \mathbb{P}(\mathcal{B}^c)\,\Delta_f(\mathcal{B}^c)$$

$$C \triangleq \sqrt{\frac{2}{N}\log\frac{2}{\delta}}$$

$$\Delta_f(\mathcal{B}) \triangleq M_f(\mathcal{B}) - m_f(\mathcal{B})$$



# Planning
## Challenges

- Non-parametric beliefs
- Belief dependent rewards
- High dimensional problems

# Planning

- How do these bounds apply to planning?
- Value function and expected reward:

$$V^{\pi}(b_k) = \underset{\boldsymbol{Z}_{k+1}}{\mathbb{E}}\left[\rho\left(b_k, \pi_k(b_k)\right)\right] + \gamma \underset{\boldsymbol{Z}_{k+1}}{\mathbb{E}}\left[V^{\pi}(b_{k+1})\right]$$

$$\underset{\boldsymbol{X}\sim b}{\mathbb{E}}\left[R(b(\boldsymbol{X}), \boldsymbol{X}, \pi(b))\right]$$

- Both have expectations we can calculate partially
- Can be done to the state or observation space, depending on our choice

# Reward Bounds

- Directly applying our primary bounds:

$$\underset{Z'}{\mathbb{E}}\left[\rho\left(b,\pi,b'\right)\right]-\mathcal{E}_{\mathcal{B}}\left[\rho\left(b,\pi,b'\right)\right]\leq\mathbb{P}\left(\mathcal{B}^{c}\right)\sup_{Z'\in\mathcal{B}^{c}}\rho\left(b,\pi,b'\right)$$

- Just like we would bound any function
- The above is with respect to the **observation space**

# Value Function Bounds

- Value function bounds:

$$\mathcal{LB}^{\pi}(b_k) \leq V^{\pi}(b_k) - \bar{V}^{\pi}(b_k) \leq \mathcal{UB}^{\pi}(b_k)$$

- We immediately arrive a cumulative reward bound:

$$\mathcal{LB}^{\pi}(b_k) = \sum_{l=k}^{k+L-1} \gamma^{l-k} \mathbb{P}\left(\mathcal{B}^{c}_{:l+1} \mid b_k, \pi\right) \inf_{Z_{:l+1} \in \mathcal{B}^{c}_{:l+1}} \rho\left(b_l, \pi_l, b_{l+1}\right)$$

Infimum with respect to all past observations, opaque

# Value Function Bounds

- Value function bounds:

$$\mathcal{LB}^{\pi}(b_k) \leq V^{\pi}(b_k) - \bar{V}^{\pi}(b_k) \leq \mathcal{UB}^{\pi}(b_k)$$

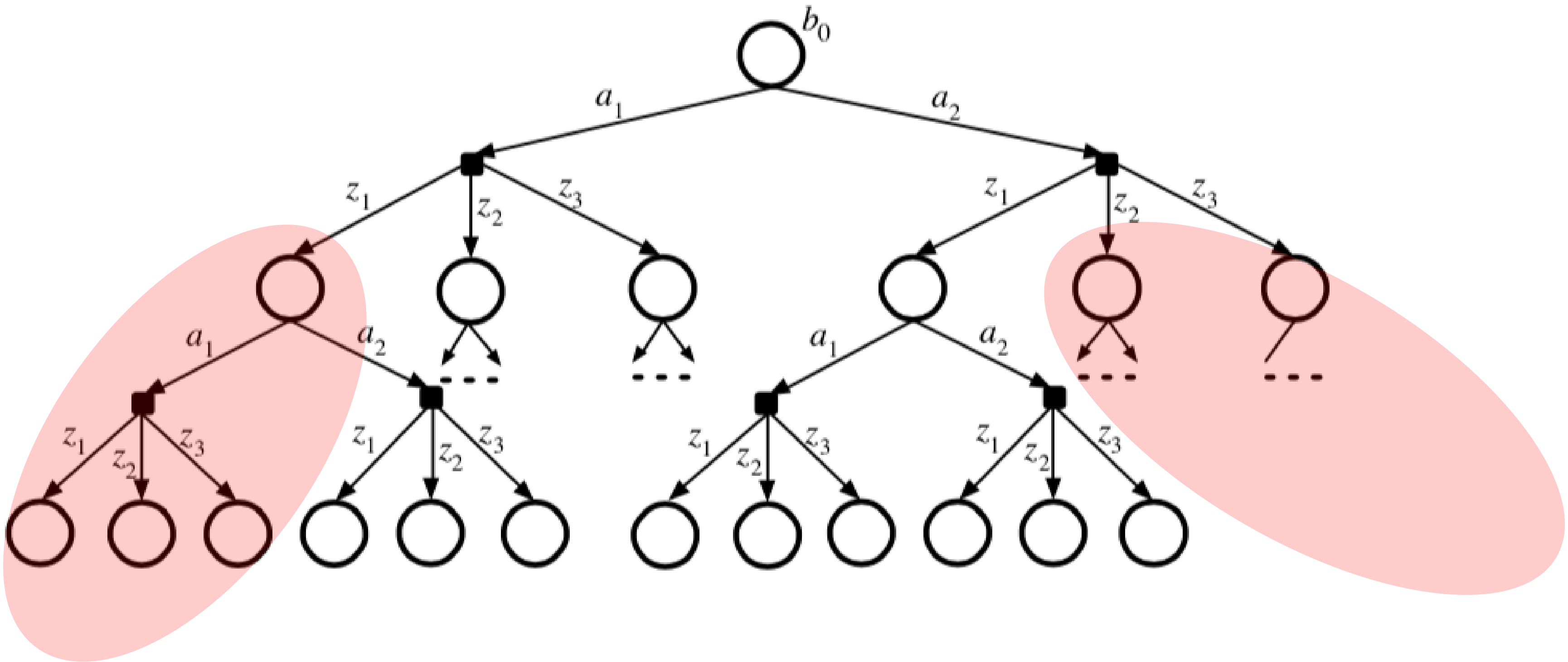
- A recursive form:

$$\mathcal{LB}^{\pi}(b_t) = \mathbb{P}\left(\mathcal{B}^{c}_{t+1} \mid b_t, \pi\right)\left(\inf_{\mathcal{B}^{c}_{t+1}} \rho(b_t, \pi_t, b_{t+1}) + \gamma \inf_{\mathcal{B}^{c}_{t+1}} \bar{V}^{\pi}(b_{t+1})\right)$$
$$+ \gamma\left(\mathbb{P}\left(\mathcal{B}^{c}_{t+1} \mid b_t, \pi\right) \inf_{\mathcal{B}^{c}_{t+1}} \mathcal{LB}(b_{t+1}) + \mathcal{E}_{\mathcal{B}_{t+1}|b_t,\pi}\left[\mathcal{LB}(b_{t+1})\right]\right)$$

Still a bit opaque, needs to be calculated online

# Value Function Bounds

- Complete branch pruning



# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer’s estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Value Function Bounds

- Value function bounds:

$$\mathcal{LB}^{\pi}(b_k) \leq V^{\pi}(b_k) - \bar{V}^{\pi}(b_k) \leq \mathcal{UB}^{\pi}(b_k)$$

- An easier recursive form:

$$\mathcal{LB}(b_t) = \mathbb{P}(\mathcal{B}^{c}_{t+1}) \inf_{\mathcal{B}^{c}_{t+1}} \rho(b_t, \pi_t, b_{t+1}) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{z}_{t+1}} [\mathcal{LB}(b_{t+1})]$$

$$\bar{V}^{\pi}(b_t) = \mathcal{E}_{\mathcal{B}_{t+1}} [\rho(b_t, \pi_t, b_{t+1})] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{z}_{t+1}} [\bar{V}^{\pi}(b_{t+1})]$$

- The tree is **unchanged**

# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer's estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Conditional Entropy

- Assumes actions are **independent** of observations
- Given with respect to the **observation space**

$$\mathcal{LB} \leq \mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) - \bar{\mathcal{H}}_Z\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) \leq \mathcal{UB}$$

- Assumes model infimums greater than zero

# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer's estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Boer's Estimator

- A common estimator for entropy is given by the Boer's estimator[1]

$$\hat{\mathcal{H}}\left(\boldsymbol{X}'\right) = \log\left(\sum_{i=1}^{N} w^i \mathrm{P}\left(Z' \mid X'^i\right)\right) - \sum_{i=1}^{N} w'^i \log\left(\mathrm{P}\left(Z' \mid X'^i\right) \sum_{j=1}^{N} w^j \mathrm{P}\left(X'^i \mid X^j\right)\right)$$

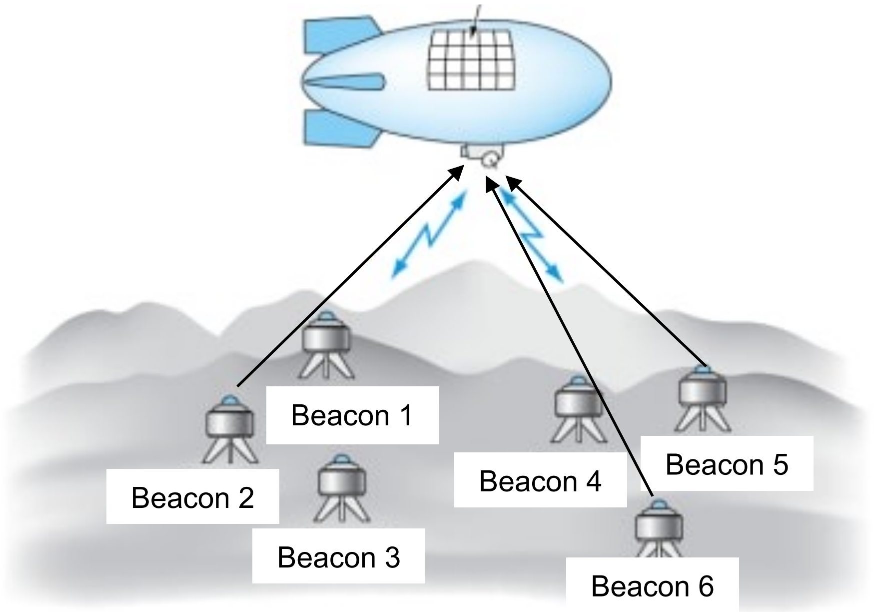
- Prior works have provided bounds with $\boldsymbol{O}(\boldsymbol{nN})^2, n \leq N$
- We provide $\boldsymbol{O}\left(\boldsymbol{n}^{2}\right)$ bounds: $\mathcal{LB} \leq \hat{\mathcal{H}}\left(\boldsymbol{X}'\right) - \bar{\mathcal{H}}\left(\boldsymbol{X}'\right) \leq \mathcal{UB}$
- Simplification is with respect to the state space
- Assumes that the model infimums are greater than zero

1) Y. Boers, H. Driessen, A. Bagchi, and P. Mandal. Particle filter based entropy. In 2010 13th International Conference on Information Fusion, pages 1–8, 2010.

2) Sztyglic, O., Indelman, V.: Speeding up online pomdp planning via simplification. In: IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS) (2022)

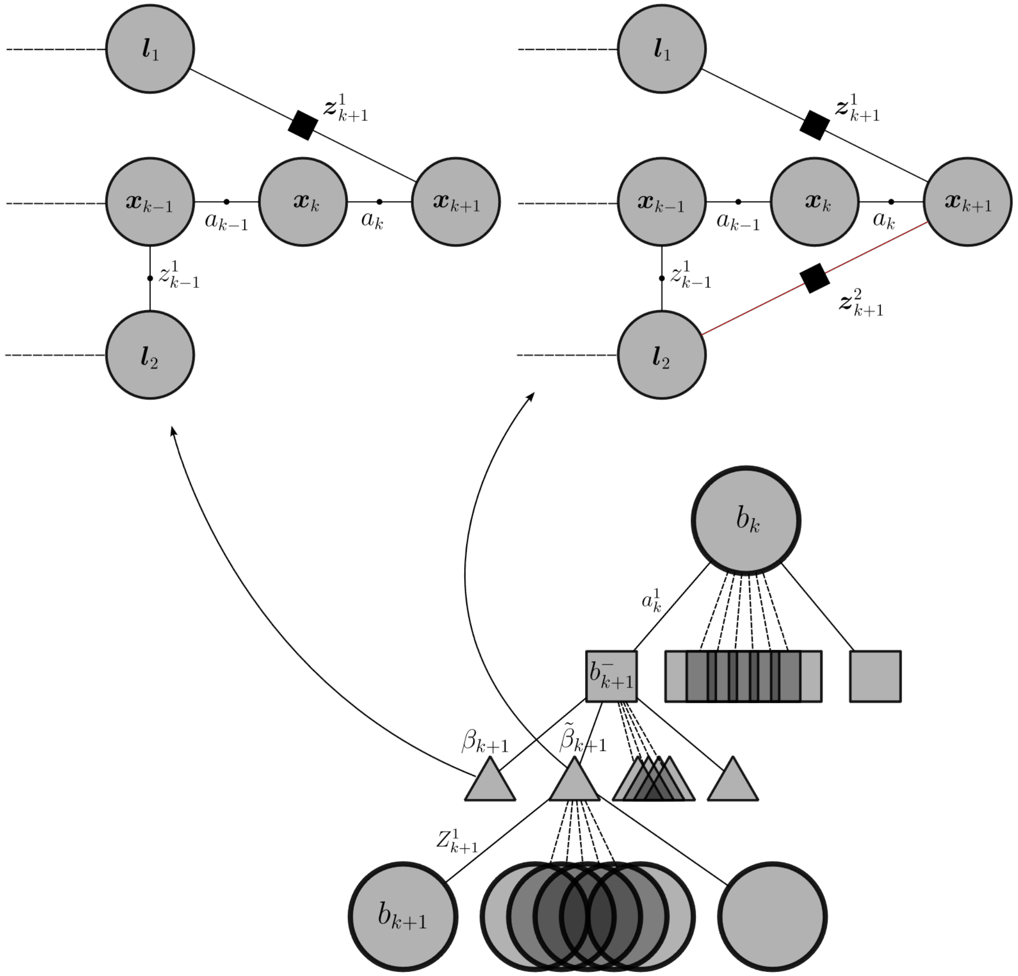
# Data Association

- Without reasoning about data association, **observation dimensionality can change**
- Multiple landmarks may be observed at a given time
- We must **reason which landmarks are observed**
- Captured via the binary random vector $\boldsymbol{\beta}$
- Example: Give the 6 landmarks available, only landmarks 2, 5, and 6 return an observation, thus $\beta = [0\ 1\ 0\ 0\ 1\ 1]^T$



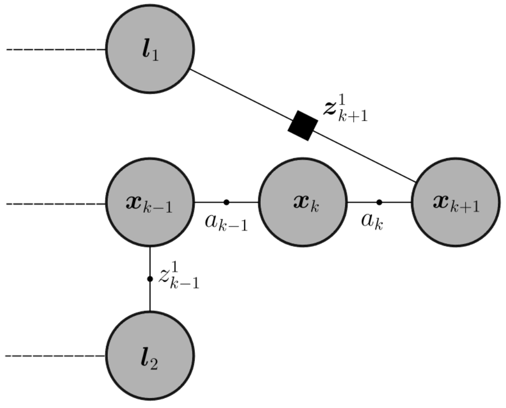
# Topology

- Many problems can be described using a factor graph (a bipartite graph representing the relationship between the variables)
- Commonly seen in **high dimensional problems** (e.g. SLAM)
- Given a realization of data association, a specific factor graph topology is specified
- We augment the traditional belief tree with the **addition of DA nodes** (triangles)
- Resulting topologies are not true factor graphs at this point as the **observations are still random** (indicated by the filled squares)

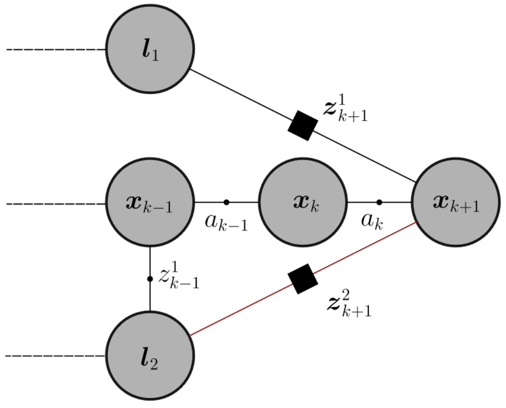


# Topology Elimination

- Similar topologies share many of the same variables.
- We reason that we can simplify calculations of one topology by utilizing a neighboring topology




Reference topology



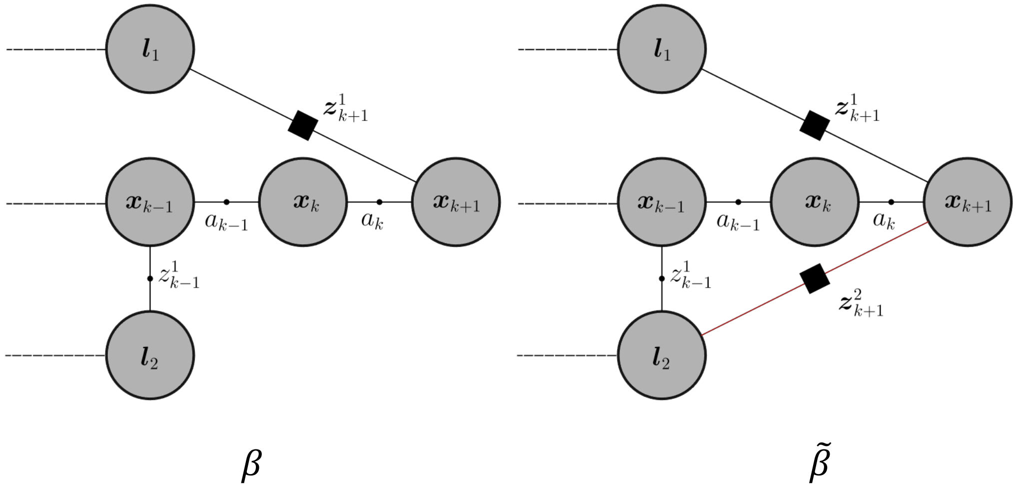

Original topology

# Contributions

- Partial expectation bounds
- Hoeffding like bounds
- Value function bounds via reward simplification
- Conditional entropy bounds
- Boer’s estimator bounds with greater computational efficiency than previous works
- Reuse of information between similar topologies when calculating conditional entropy

# Conditional Entropy for Data Association (DA)

- Bound the conditional entropy of a specific DA realization ($\beta$) with respect to data from a similar DA realization ($\tilde{\beta}$).




$$\mathcal{LB}(\beta_{\text{diff}}) \leq \mathcal{H}\left(X \mid Z, \tilde{\beta}\right) - \bar{\mathcal{H}}\left(X \mid Z, \beta, \tilde{\beta}\right) \leq \mathcal{UB}(\beta, \tilde{\beta})$$

# Q-Function Bounds

- Utilizing the previous bounds as bound on the expected reward when the reward is entropy

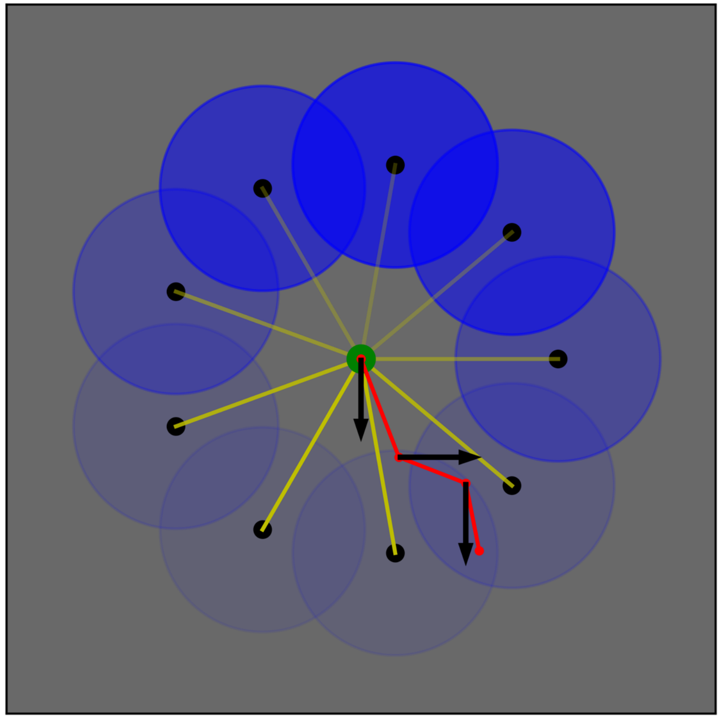
$$\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\left[\mathop{\mathbb{E}}_{\boldsymbol{Z}|\boldsymbol{\beta}}\left[\rho\left(b\right)\right]\right] - \mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{Z}|\boldsymbol{\beta}}\left[\rho\left(b\right)\right]\right] = \sum_{\beta_i \in \mathcal{B}^c} \mathbb{P}\left(\beta_i\right) \mathop{\mathbb{E}}_{\boldsymbol{Z}|\beta_i}\left[\rho\left(b\right)\right]$$

$$\bar{V}^{\pi}\left(b_t\right) = \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\bar{V}^{\pi}\left(b_{t+1}\right)\right]$$

- Utilizing over the planning horizon allows for bounds on the Q-function
- Can be used with Bellman optimality to:
  - Provide bounds on the optimal value function
  - If the bounds are tight enough can be used for optimal action selection

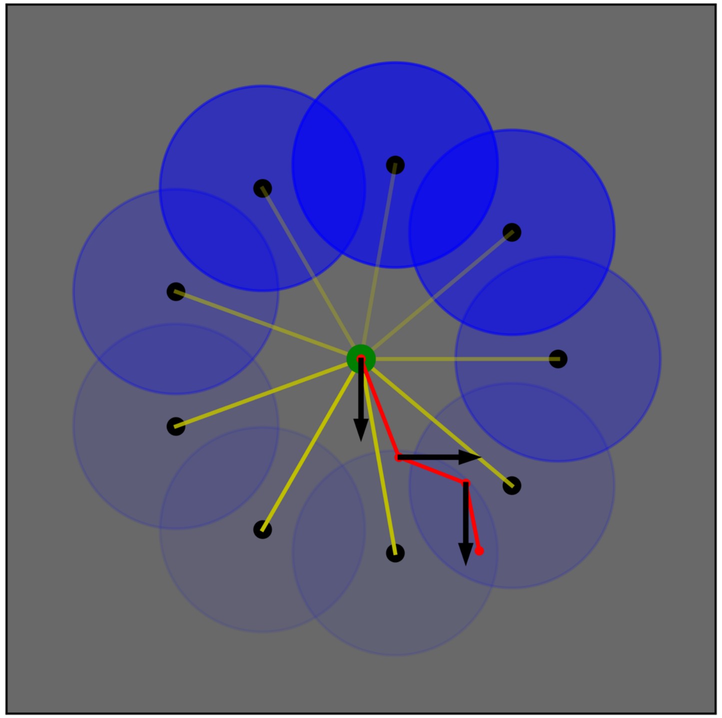
# Scenario

- Information gathering task in $R^2$ with 9 landmarks
- Landmarks are uncertain, part of the belief
- Planning horizon of 1
- Action space = cardinal directions
- Smoothing problem, past poses are added to state
- Simplification is performed with respect to the DA space
- We bound an estimator of the conditional entropy



# Scenario

- Green is starting point
- Red line is the ground truth trajectory
- Blue intensity indicates landmark success probability
- Yellow intensity indicates prior landmark certainty
- Black arrows are $argmax_a\ estimator\ lower\ bound$

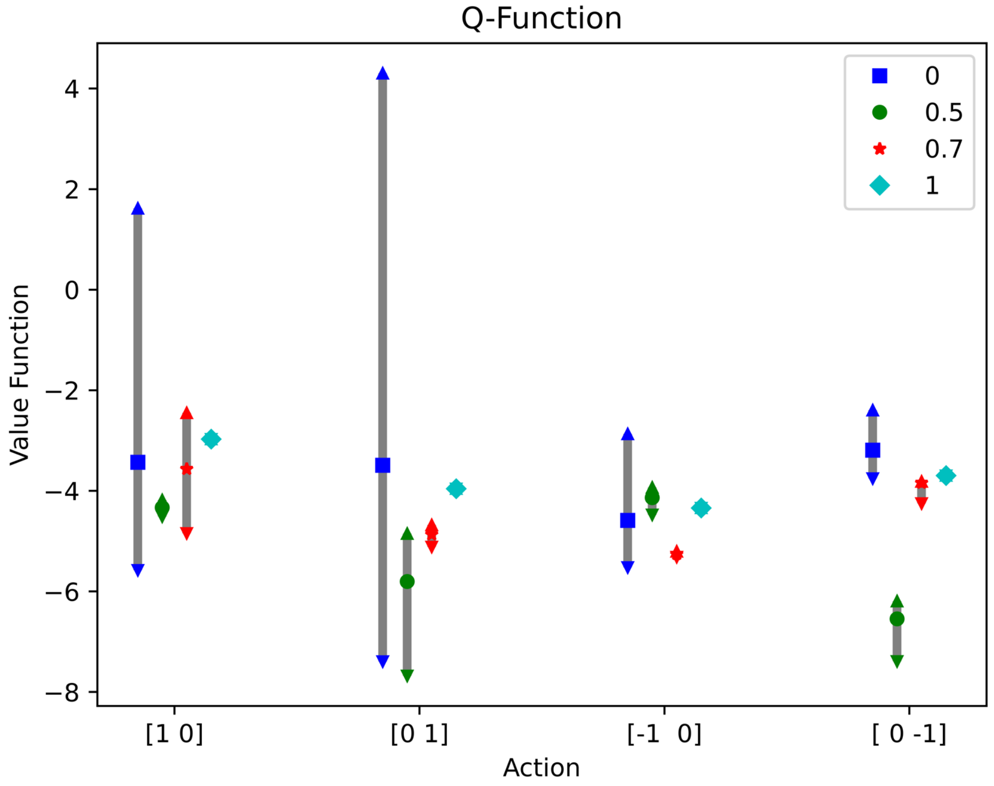


# Results

- $\kappa$: approximate fraction of DA space that is calculated
- $\kappa$ **directly proportional** to runtime
- $\kappa$ **directly proportional** to bound tightness
- Each $\kappa$ run is calculated from different particles resulting in a different estimator

**Table 1:** A value of $q = 2$ was selected for corollary 10. 150 samples were used for inference, 150 observations per action for sparse sampling, and 100 samples for reward calculations.

| $\kappa^3$ | No. Eliminated Factors | Reward Runtime [s] | Bounds Runtime [s] | Speedup |
|---|---|---|---|---|
| 1 | 0 | $876.5 \pm 164.5$ | $874.0 \pm 164.5$ | 1.0 |
| 0.7 | $88 \pm 45$ | $991.1 \pm 446.6$ | $743.7 \pm 283.4$ | $1.3 \pm 0.1$ |
| 0.5 | $189 \pm 92$ | $720.8 \pm 43.0$ | $295.0 \pm 94.5$ | $2.6 \pm 0.4$ |
| 0 | $330 \pm 164$ | $651.7 \pm 123.5$ | $16.3 \pm 0.6$ | $39.8 \pm 7.1$ |



# Open Questions

- How to select the best subset of a given size for the tightest bounds
- How to efficiently select a subset for online use
- How to select $\beta_{reff}$
- Implementation of a full adaptive planning algorithm than can select the optimal action
- Incorporation of our Hoeffding like bounds into planning

# Summary

- Generalization of Markov inequality
- Generalization of Hoeffding’s inequality
- Simplification of rewards and value functions for planning
- Bounds on the conditional entropy via observation space simplification
- Bounds on the Boers estimator via state space simplification
- Conditional entropy bounds for problems with topology and data association