Planning under Uncertainty using Distributions over Posteriors

Nicholas Roy
Joint work with Ruijie He, Emma Brunskill

Autonomous Micro Air Vehicle Flight Indoors
Robust Robotics Group
CSAIL, MIT
Planning under Uncertainty using Distributions over Posteriors

Nicholas Roy
Joint work with Ruijie He, Emma Brunskill

Autonomous Micro Air Vehicle Flight Indoors
Robust Robotics Group
CSAIL, MIT
Joint work with Abraham Bachrach, Ruijie He, Sam Prentice

Autonomous Entry
Technical Challenges

- Very fast flight dynamics
- Limited or no external positioning system
  Fast, accurate sensor processing is essential

- Limited on-board sensing
- Limited ability to process sensor data fast enough
  Planner must take into account uncertainty from sensor limitations

- Limited prior knowledge
- Limited ability to compute complex plans
  Need efficient solutions to complex planning problems
Technical Challenges

- Very fast flight dynamics
- Limited or no external positioning system

Fast, accurate sensor processing is essential

- Limited on-board sensing
- Limited ability to process sensor data fast enough

Planner must take into account uncertainty from sensor limitations

- Limited prior knowledge
- Limited ability to compute complex plans

Need efficient solutions to complex planning problems
Traditional Forward Search

Planning phase
– Create AND-OR tree
– Fringe nodes approx. value function
– Estimate current belief value
– Choose best policy

Execution phase
– Executes policy, updates belief

Repeat Cycle
Traditional Forward Search

Planning phase
- Create AND-OR tree
- Fringe nodes approx. value function
- Estimate current belief value
- Choose best policy

Execution phase
- Executes policy, updates belief

Repeat Cycle
Traditional Forward Search

Planning phase
- Create AND-OR tree
- Fringe nodes approx. value function
- Estimate current belief value
- Choose best policy

Execution phase
- Executes policy, updates belief

Repeat Cycle
Traditional Forward Search

Planning phase
- Create AND-OR tree
- Fringe nodes approx. value function
- Estimate current belief value
- Choose best policy

Execution phase
- Executes policy, updates belief

Repeat Cycle
Traditional Forward Search

Advantages
- Focus on reachable beliefs
- Leverage factored models
- Applicable to much wider range of model types (not just LQG or discrete POMDPs)

Challenges
- Scales poorly with horizon length
- $O(|A||Z|)^H$

Hypothesis
- Conditioning on the observation after every action is unnecessary for many tasks
Traditional Forward Search

**Advantages**
- Focus on reachable beliefs
- Leverage factored models
- Applicable to much wider range of model types (not just LQG or discrete POMDPs)

**Challenges**
- Scales poorly with horizon length
- $O(|A||Z|)^H$

---

**Hypothesis**

- Conditioning on the observation after *every* action is unnecessary for many tasks
Faster Forward Search

**Macro-actions**
- Fixed-length, open-loop policies

- Restricts policy class
- Longer horizon-search

Plan using this as a single action
Faster Forward Search

Macro-actions
- Fixed-length, open-loop policies

First action of Macro-action

Second action of Macro-action

- Restricts policy class
- Longer horizon-search

Plan using this as a single action
Forward Search with Macro-actions

Challenges
- How to compute expected reward?
- Just another expectation, but over observations

Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences
Forward Search with Macro-actions

Challenges
- How to compute expected reward?
- Just another expectation, but over observations

Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences
Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences

2. Sample from possible observation sequences

First action of Macro-action
Sample observation

Second action of Macro-action
Sample observation

Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences

2. Sample from possible observation sequences

3. Compute distribution over beliefs analytically
Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences
2. Sample from possible observation sequences

---

First action of Macro-action
Sample observation

Second action of Macro-action
Sample observation

---

Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences
2. Sample from possible observation sequences
3. Compute distribution over beliefs analytically
Analytic distribution over posterior beliefs (PBD)

- Gaussian beliefs
  - Linear-Gaussian models
  - Kalman filter
- Approximate generalizations exist for non-linear-Gaussian models

Forward Search with Macro-actions

- Never branch on received observations
- Long, open-loop plan
- Chained macro-actions
- Expected value guaranteed to be lower bound of optimal value
  - If reward function is weighted sum of Gaussians or weighted sum of polynomials

Open-loop: PBDE
Analytic distribution over posterior beliefs (PBD)

- Gaussian beliefs
  - Linear-Gaussian models
  - Kalman filter
- Approximate generalizations exist for non-linear-Gaussian models

Forward Search with Macro-actions

- Never branch on received observations
- Long, open-loop plan
- Chained macro-actions
- Expected value guaranteed to be lower bound of optimal value
  - If reward function is weighted sum of Gaussians or weighted sum of polynomials

Open-loop: PBDE
Conditioning *sometimes*

After each macro action
- Sample from posterior belief distribution
- Compute best action per sample
- Analytic calculation of expected reward no longer possible

Samples from distribution over beliefs

**Forward Search with Macro-Actions**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Computation Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBDE</td>
<td>$O(M^{H/L}LD^3W)$</td>
</tr>
<tr>
<td>PBD</td>
<td>$O(M^{H/L}N_s^{(H/L)-1}LD^3W)$</td>
</tr>
<tr>
<td>Discrete-state, full</td>
<td>$O(M^{H/L}N_s^{(H/L)}Lg^{2D})$</td>
</tr>
<tr>
<td>No macro actions</td>
<td>$O(</td>
</tr>
</tbody>
</table>

$D$ = # state dimensions, $M$ = # macro-actions, $L$ = length of macro-action, $N_s$ = # samples, $W$ = constant, $g$ = # discrete states/dimension
Conditioning *sometimes*

After each macro action
- Sample from posterior belief distribution
- Compute best action per sample
- Analytic calculation of expected reward no longer possible

**Forward Search with Macro-Actions**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Computation Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBDE</td>
<td>$O(M^{H/L} LD^3 W)$</td>
</tr>
<tr>
<td>PBD</td>
<td>$O(M^{H/L} N_s^{(H/L)-1} LD^3 W)$</td>
</tr>
<tr>
<td>Discrete-state, full</td>
<td>$O(M^{H/L} N_s^{(H/L)} L g^{2D})$</td>
</tr>
<tr>
<td>No macro actions</td>
<td>$O(</td>
</tr>
</tbody>
</table>

$D= \# \text{state dimensions}, M= \# \text{macro-actions}, L= \text{length of macro-action}, N_s=\# \text{samples}, W=\text{constant}, g= \# \text{discrete states/dimension}$
Rocks sample

Derived from RockSample (Smith & Simmons 2004)

Information Rocks sample

Derived from RockSample (Smith & Simmons 2004)
Rocks Sample

Information Rocks Sample

Derived from RockSample (Smith & Simmons 2004)
Information Rocksample

SARSOP
(Kurniawati et al. 2008)

PBD
Conditional macro-actions

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ave. rewards</th>
<th>Online time(s)</th>
<th>Offline time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMDP</td>
<td>1.11±0.43</td>
<td>0.01</td>
<td>3.03</td>
</tr>
<tr>
<td>HSVI</td>
<td>6.78±2.46</td>
<td>0.051</td>
<td>1000</td>
</tr>
<tr>
<td>SARSOP</td>
<td>8.46±2.46</td>
<td>0.07</td>
<td>25000</td>
</tr>
<tr>
<td>RTBSS</td>
<td>9.78±1.69</td>
<td>17.64</td>
<td>0</td>
</tr>
<tr>
<td>MAC</td>
<td>13.68±1.86</td>
<td>15.39</td>
<td>0</td>
</tr>
<tr>
<td>PBD</td>
<td>14.49±1.73</td>
<td>1.26</td>
<td>0</td>
</tr>
<tr>
<td>MAD</td>
<td>15.88±1.58</td>
<td>4.81</td>
<td>0</td>
</tr>
<tr>
<td>Fully obs.</td>
<td>21.37</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>
Information Rocksample

Algorithm | Ave. rewards | Online time(s) | Offline time (s)
--- | --- | --- | ---
QMDP | 1.11±0.43 | 0.01 | 3.03
HSVI | 6.78 ±2.46 | 0.051 | 1000
SARSOP | 8.46 ±2.46 | 0.07 | 25000
RTBSS | 9.78 ±1.69 | 17.64 | 0
MAC | 13.68 ± 1.86 | 15.39 | 0
PBD | 14.49 ±1.73 | 1.26 | 0
MAD | 15.88 ± 1.58 | 4.81 | 0
Fully obs. | 21.37 | N.A. | N.A.

(SARSOP (Kurniawati et al. 2008))

Conditional macro-actions
Information Rocks Sample

Scales to much larger problems

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ave. rewards</th>
<th>Online time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>42.64</td>
<td>310.05</td>
</tr>
<tr>
<td>PBD</td>
<td>43.68</td>
<td>60.81</td>
</tr>
<tr>
<td>MAD</td>
<td>51.70</td>
<td>101.92</td>
</tr>
<tr>
<td>Fully obs.</td>
<td>66.61</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Target Monitoring

Reward function
- Rewards for correctly reporting target in regions
- Penalty for incorrect report

Observations
- Limited, noisy observations of targets
- Field-of-view & quality height-dependent
Information Rocks Sample

Scales to much larger problems

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Ave. rewards</th>
<th>Online time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>42.64</td>
<td>310.05</td>
</tr>
<tr>
<td>PBD</td>
<td>43.68</td>
<td>60.81</td>
</tr>
<tr>
<td>MAD</td>
<td>51.70</td>
<td>101.92</td>
</tr>
<tr>
<td>Fully obs.</td>
<td>66.61</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Target Monitoring

**Reward function**
- Rewards for correctly reporting target in regions
- Penalty for incorrect report

**Observations**
- Limited, noisy observations of targets
- Field-of-view & quality height-dependent
Quantitative Results

<table>
<thead>
<tr>
<th></th>
<th>Targets reported</th>
<th>Ground truth</th>
<th>Flight time (s)</th>
<th>Dist. traveled (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-Single</td>
<td>1</td>
<td>7</td>
<td>484.15</td>
<td>243.36</td>
</tr>
<tr>
<td>NBO</td>
<td>1</td>
<td>4</td>
<td>435.25</td>
<td>247.01</td>
</tr>
<tr>
<td>PBD</td>
<td>4</td>
<td>6</td>
<td>474.64</td>
<td>282.51</td>
</tr>
</tbody>
</table>

**WT-Single**
- Go to target with largest uncertainty

**NBO (Scott et al. 2009)**
- Uses Kalman filter
- Assumes most likely posterior belief after macro-action (instead of using full distribution of beliefs)
Quantitative Results

<table>
<thead>
<tr>
<th></th>
<th>Targets reported</th>
<th>Ground truth</th>
<th>Flight time (s)</th>
<th>Dist. traveled (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-Single NBO</td>
<td>1</td>
<td>7</td>
<td>484.15</td>
<td>243.36</td>
</tr>
<tr>
<td>PBD</td>
<td>4</td>
<td>6</td>
<td>474.64</td>
<td>282.51</td>
</tr>
</tbody>
</table>

**WT-Single**

- Go to target with largest uncertainty
- NBO (Scott et al. 2009)
- Uses Kalman filter
- Assumes most likely posterior belief after macro-action (instead of using full distribution of beliefs)
Automatic Macro-action Generation

Initial macro-actions provided by any controller
- e.g., MDP solution, LQG controller

Anytime Search
- Incremental refinement of macro-actions
- Reduces sensitivity to suboptimality in initial macro-action set

Summary

• Macro-actions and posterior belief distributions substantially accelerate search
  – Especially in large, high-dimensional domains

• Macro-actions provide a principled way to combine multistep controllers with higher-level planning

• Trade-off between depth of search and number of conditional outcomes examined
Automatic Macro-action Generation

Initial macro-actions provided by any controller
- e.g., MDP solution, LQG controller

Anytime Search
- Incremental refinement of macro-actions
- Reduces sensitivity to suboptimality in initial macro-action set

Summary

- Macro-actions and posterior belief distributions substantially accelerate search
  - Especially in large, high-dimensional domains

- Macro-actions provide a principled way to combine multistep controllers with higher-level planning

- Trade-off between depth of search and number of conditional outcomes examined