Planning under Uncertainty using Distributions over Posteriors

Nicholas Roy
Joint work with Ruijie He, Emma Brunskill

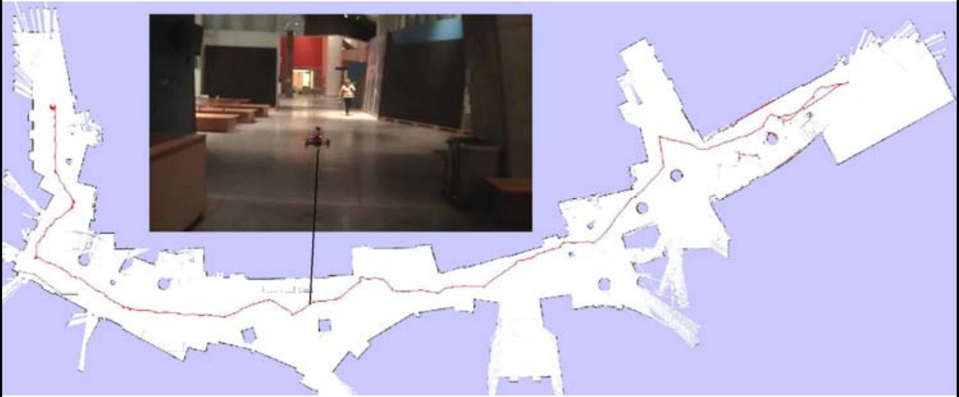
Autonomous Micro Air Vehicle Flight Indoors

Robust Robotics Group
CSAIL, MIT

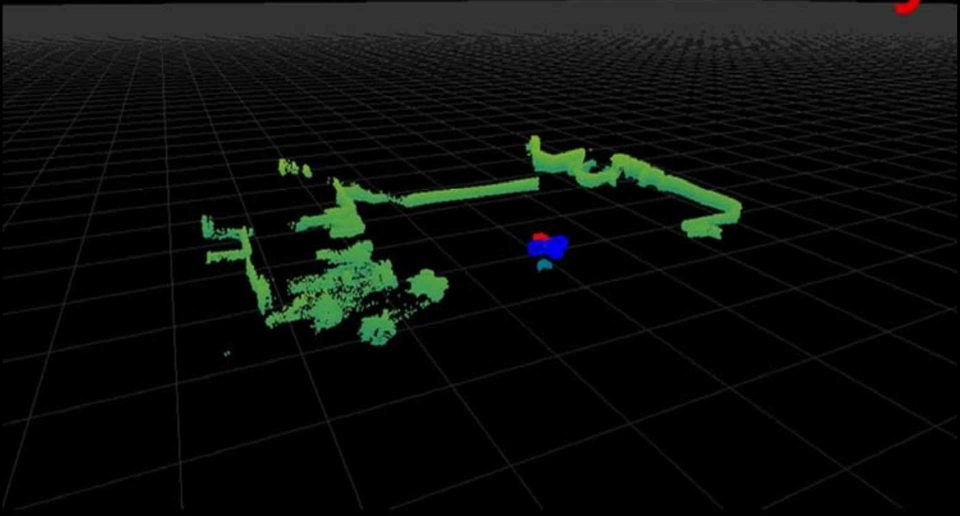
CSAIL

MIT AEROASTRO

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

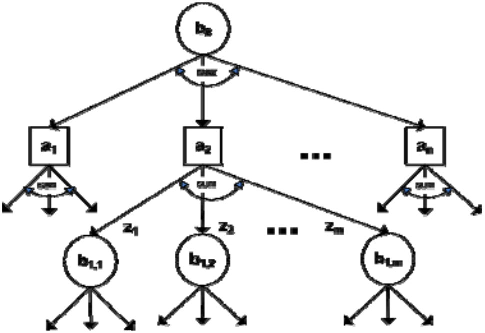
- Limited prior knowledge
- Limited ability to compute complex plans

Need efficient solutions to complex planning problems

Traditional Forward Search

Planning phase

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



Execution phase

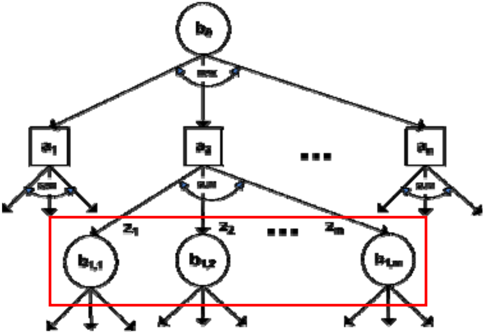
- Executes policy, updates belief

Repeat Cycle

Traditional Forward Search

Planning phase

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

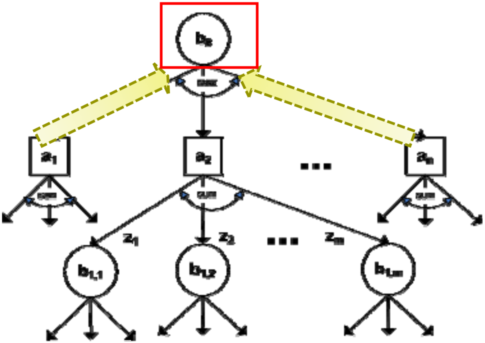


Execution phase

- Executes policy, updates belief

Repeat Cycle

Traditional Forward Search



<u>Planning phase</u>

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

<u>Execution phase</u>

- Executes policy, updates belief

<u>Repeat Cycle</u>

Traditional Forward Search

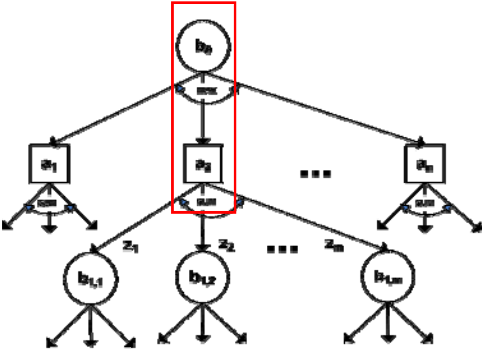
Planning phase

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

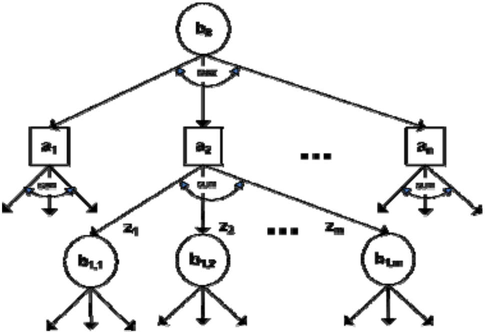
Execution phase

- 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
- $O(|A||Z|)^H$

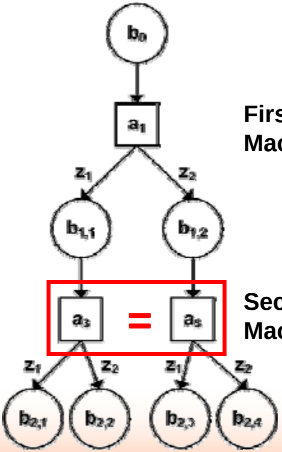
Hypothesis

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

Faster Forward Search

<u>Macro-actions</u>

– Fixed-length, open-loop policies



First action of Macro-action

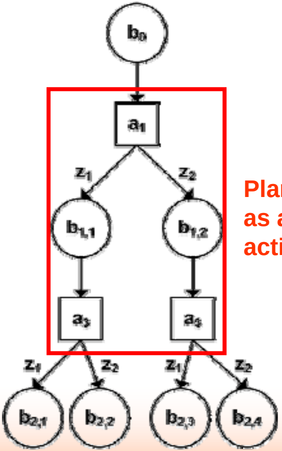
Second action of Macro-action

Faster Forward Search

Macro-actions

- Fixed-length, open-loop policies

- Restricts policy class
- Longer horizon-search

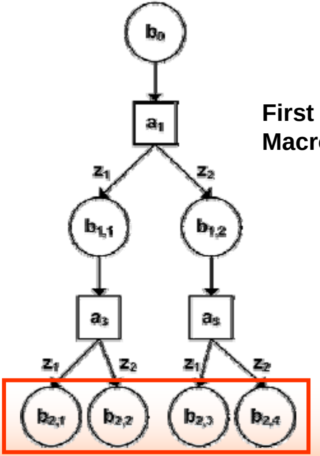


Plan using this as a single action

Forward Search with Macro-actions

Challenges

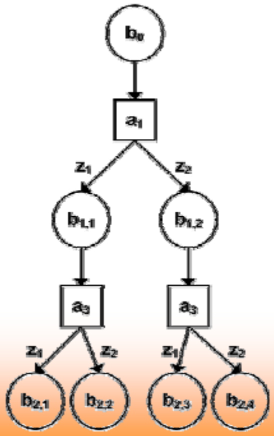
- How to compute expected reward?
- Just another expectation, but over observations



First action of Macro-action

Evaluating Macro-actions

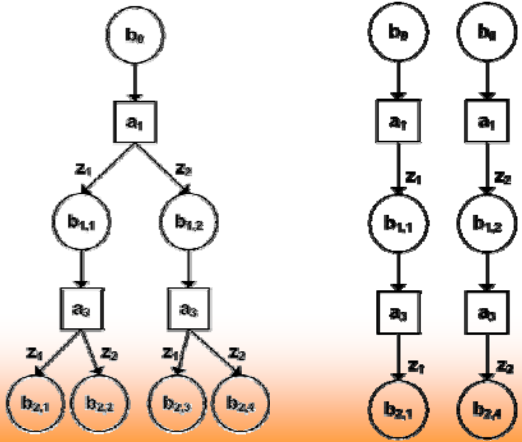
1. Exhaustively enumerate all possible observation sequences



Evaluating Macro-actions

1. Exhaustively enumerate all possible observation sequences

2. Sample from possible observation sequences



First action of Macro-action

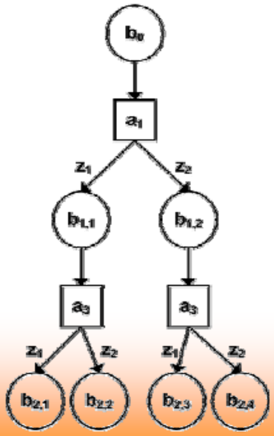
Sample observation

Second action of Macro-action

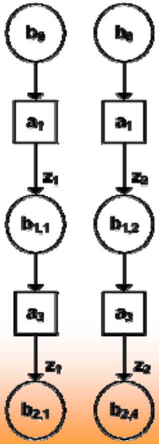
Sample observation

Evaluating Macro-actions

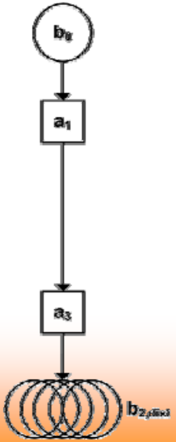
1. Exhaustively enumerate all possible observation sequences



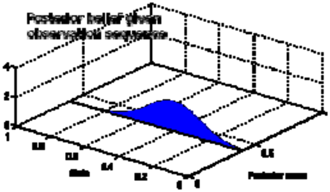
2. Sample from possible observation sequences

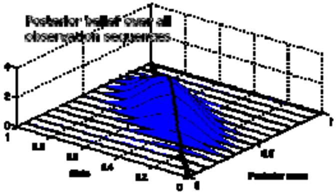


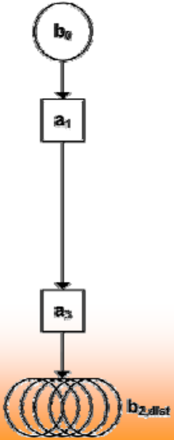
3. Compute distribution over beliefs analytically



Analytic distribution over posterior beliefs (PBD)





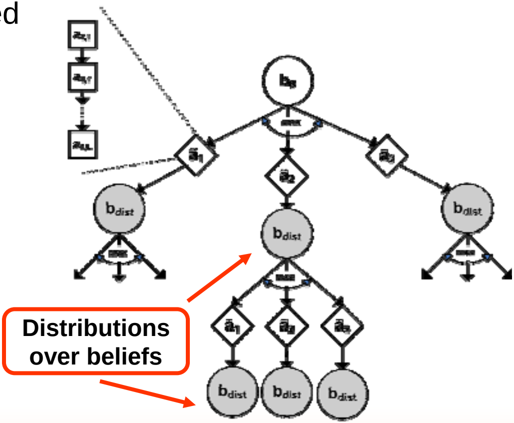


- Gaussian beliefs
  - Linear-Gaussian models
  - Kalman filter
- Approximate generalizations exist for non-linear-Gaussian models

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



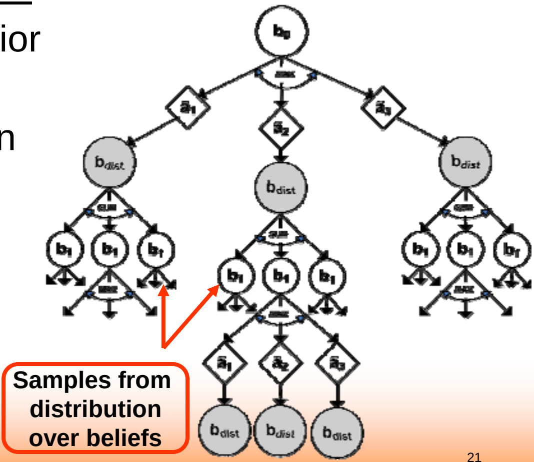
Open-loop: PBDE

Conditioning *sometimes*

After each macro action

- Sample from posterior belief distribution
- Compute best action per sample

- Analytic calculation of expected reward no longer possible

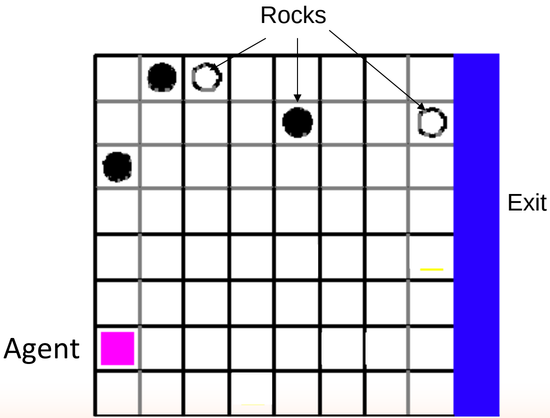


Forward Search with Macro-Actions

Algorithm	Computation Complexity
PBDE	$O(M^{H/L} L D^3 W)$
PBD	$O(M^{H/L} N_s^{(H/L)-1} L D^3 W)$
Discrete-state, full	$O(M^{H/L} N_s^{(H/L)} L g^{2D})$
No macro actions	$O(\|A\|^H \|Z\|^H)$

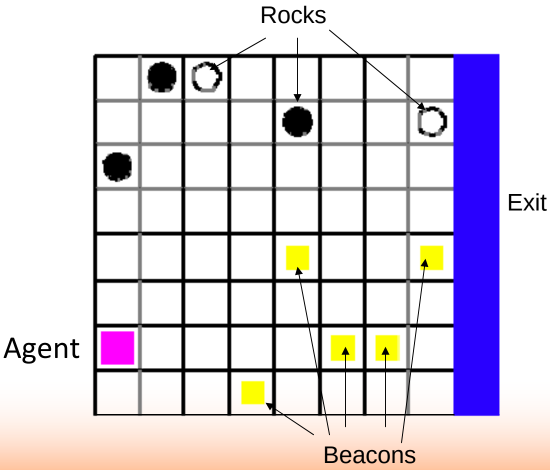
D= # state dimensions, M= # macro-actions, L= length of macro-action, N_s=# samples, W=constant, g= # discrete states/dimension

Rocksample



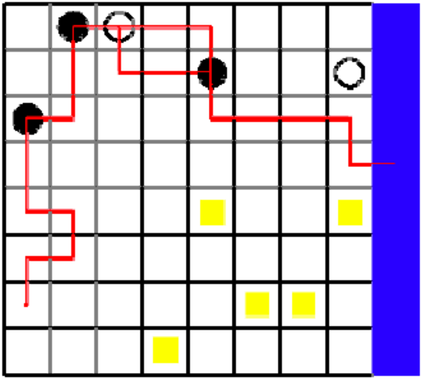
Derived from RockSample (Smith & Simmons 2004)

Information Rocksample

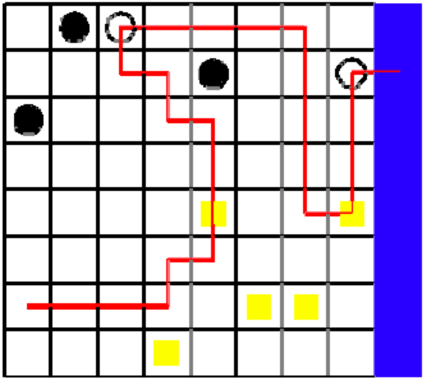


Derived from RockSample (Smith & Simmons 2004)

Information Rocksample



SARSOP
(*Kurniawati et al. 2008*)



PBD
Conditional macro-actions

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.

Information Rocksample

Scales to much larger problems

	ISRS [100,30]	
Algorithm	Ave. rewards	Online time(s)
MAC	42.64	310.05
PBD	43.68	**60.81**
MAD	**51.70**	101.92
Fully obs.	66.61	N.A.

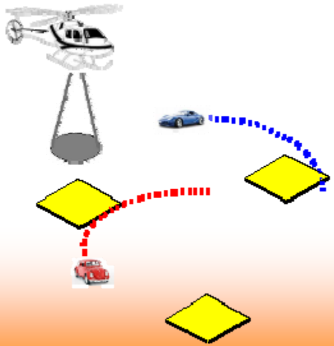
Target Monitoring

<u>Reward function</u>

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

<u>Observations</u>

- Limited, noisy observations of targets
- Field-of-view & quality height-dependent



Quantitative Results

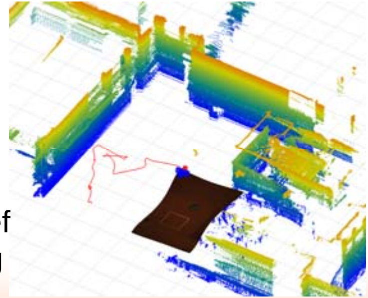
	Targets reported	Ground truth	Flight time (s)	Dist. traveled (m)
WT-Single	1	7	484.15	243.36
NBO	1	4	435.25	247.01
PBD	**4**	6	474.64	282.51

<u>WT-Single</u>

- Go to target with largest uncertainty

<u>NBO (Scott et al. 2009)</u>

- Uses Kalman filter
- Assumes most likely posterior belief after macro-action (instead of using full distribution of beliefs)



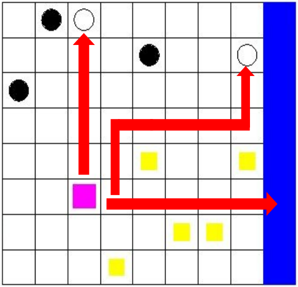
Automatic Macro-action Generation

<u>Initial macro-actions provided by any controller</u>

- e.g., MDP solution, LQG controller

<u>Anytime Search</u>

- 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