Solving Continuous-State POMDPs via Density Projection

Enlu Zhou, Michael C. Fu, and Steven I. Marcus

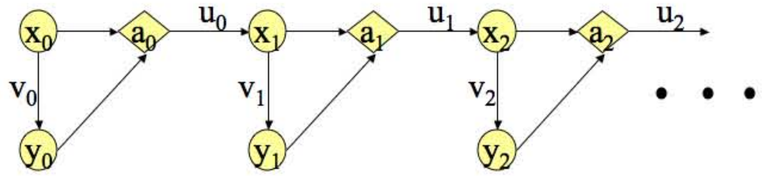
The Institute for Systems Research

University of Maryland

Introduction

Partially observable Markov decision processes (**POMDPs**) model sequential decision making under uncertainty with partially observed state information. Research on numerical solution methods for POMDPs has primarily focused on discrete-state models, and these algorithms do not generally extend to **continuous-state POMDPs**, due to the infinite dimensionality of the belief space. We develop a method for solving continuous-state POMDPs by effectively reducing the dimensionality of the belief space via density projections.

POMDP



System dynamics $x_{k+1} = f(x_k, a_k, u_k),\ k = 0, 1, \ldots.$

Observation $y_k = h(x_k, a_{k-1}, v_k),\ k = 1, 2, \ldots.$

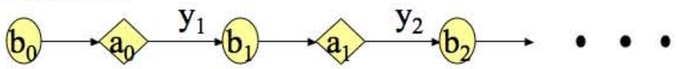
Objective function $\min J = E_{x_0, u_k, v_k, k=0,1,\ldots}\{\sum_{k=1}^{\infty} \gamma^k g(x_k, a_k)\}.$

POMDP → Belief MDP

Belief State: conditional density of the current state given all the available information

$$b_k = p(x_k \mid y_0, y_1, \ldots, y_k, a_0, a_1, \ldots, a_{k-1})$$

Belief MDP



System dynamics $b_{k+1}(x_{k+1}) = \psi(b_k, a_k, y_{k+1})$

Objective function $\min J = E_{x_0, u_k, v_k, k=0,1,\ldots}\{\sum_{k=1}^{\infty} \gamma^k \tilde{g}(b_k, a_k)\}.$

Belief MDP → Projected Belief MDP

Orthogonal density projection is defined as

$$\text{projection of } q \text{ on } \Pi = \arg\min_{p \in \Pi} D_{KL}(q \,\|\, p)$$

$$D_{KL}(q \,\|\, p) = \int_{-\infty}^{+\infty} q(x) \log \frac{q(x)}{p(x)} dx$$

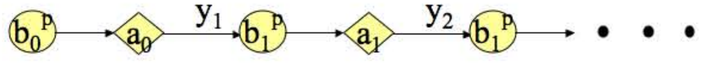
For **an exponential family of densities**

$$\Pi = \{ p(x,\theta) = \exp[\theta^T c(x) - \varphi(\theta)],\ \theta \in \Theta \},$$

it can be carried out analytically

$$E_q[c(X)] = E_{p(\cdot,\theta)}[c(X)].$$

Using the orthogonal density projection, the original Belief MDP can be converted to a **Projected Belief MDP** by projecting the belief states and the cost function.

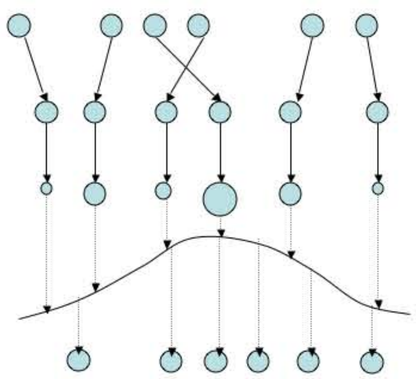


System dynamics $b^p_{k+1}(x_{k+1}) = \psi(b^p_k, a_k, y_{k+1})^p$

Objective function $\min J = E_{x_0, u_k, v_k, k=0,1,\ldots}\{\sum_{k=1}^{\infty} \gamma^k \tilde{g}(b^p_k, a_k)\}.$

The original Belief MDP is an **infinite-dimensional** continuous-state MDP, whereas the Projected Belief MDP is a **low-dimensional** continuous-state MDP and thus can be solved by numerous methods.

Projection Particle Filtering



Propagation (Importance sampling)

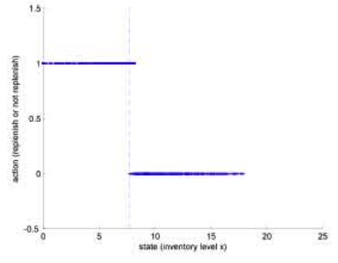
Updating (Bayes' rule)

Projection

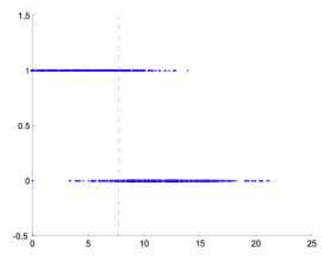
Resampling

Example: Inventory Control1

Obsv. Noise std. dev. = 0.1



Obsv. Noise std. dev. = 3.1



The actions taken by our algorithm v.s. the actual inventory levels. The dotted vertical line is the optimal threshold policy under full observation. Our method picks actions very close to the optimal policy under full observation when the observation noise is small (left fig.). Larger observation means less information (right fig.).

Obsv. stdev	Our Method	CE policy	Greedy policy
0.1	12.849 (0.12%)	12.842 (0.06%)	25.454 (98.34%)
0.5	12.864 (0.23%)	12.867 (0.26%)	25.457 (98.36%)
0.9	12.904 (0.55%)	12.908 (0.57%)	25.450 (98.30%)
1.3	12.973 (1.08%)	12.977 (1.12%)	25.356 (97.57%)
1.7	13.066 (1.81%)	13.100 (2.07%)	25.324 (97.32%)
2.1	13.123 (2.25%)	13.183 (2.72%)	25.332 (97.38%)
2.5	13.250 (3.24%)	13.314 (3.74%)	25.402 (97.92%)
2.9	13.374 (4.21%)	13.458 (4.86%)	25.478 (98.52%)
3.1	13.444 (4.75%)	13.527 (5.40%)	25.553 (99.10%)

Each entry shows the average cost, and in the parentheses the percentage error from the average cost under full observation using the optimal threshold policy.