Planning in Factored Hybrid-MDPs

Milos Hauskrecht
Department of Computer Science, University of Pittsburgh

Branislav Kveton
Intelligent Systems Program, University of Pittsburgh

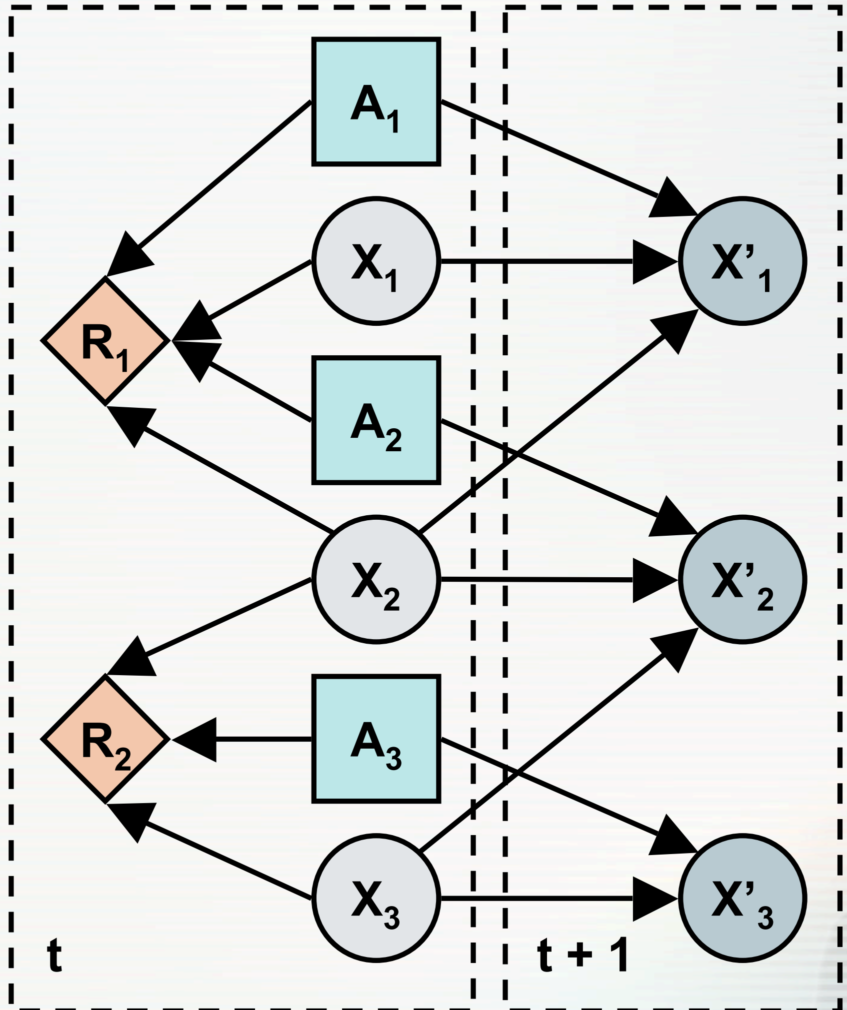
Carlos Guestrin
Department of Computer Science, CMU

Factored MDPs

- **Limitations**
  - Standard MDP formulation permits only flat discrete state and action spaces
  - Real-world problems are complex, factored, and involve continuous quantities
- **Factored MDPs with discrete components**
  - State and action spaces grow exponentially in the number of variables
  - Traditional methods for solving MDPs are polynomial in the size of state and action spaces
- **Factored MDPs with continuous components**
  - State and action spaces are infinitely large
  - optimal value function or policy may not have finite support
  - Naive discretization of continuous variables often leads to an exponential complexity of solution methods

Hybrid Factored MDPs

- A **hybrid factored MDP (HMDP)** is a 4-tuple M = (X, A, P, R):
  - **X** = (X_D, X_C) is a vector of state variables (discrete or continuous)
  - **A** = (A_D, A_C) is a vector of action variables (discrete or continuous)
  - **P** and **R** are factored transition and reward models represented by a dynamic Bayesian network (DBN)



Optimal Value Function and Policy

- **Optimal policy π^* maximizes the infinite horizon discounted reward:**

$$E_\pi\left[\sum_{t=0}^{\infty} \gamma^t R(\mathbf{x}_t, \pi(\mathbf{x}_t))\right]$$

- **Optimal value function V^* is a fixed point of the Bellman equation:**

$$V^*(\mathbf{x}) = \sup_{\mathbf{a}}\left[R(\mathbf{x},\mathbf{a}) + \gamma \sum_{\mathbf{x}_D} \int_{\mathbf{x}'_C} P(\mathbf{x}' \mid \mathbf{x},\mathbf{a}) V^*(\mathbf{x}')\, d\mathbf{x}'_C\right]$$

- **Linear value function approximation**
  - **A compact representation of an MDP does not guarantee the same structure in the optimal value function or policy**
  - **Approximation by a linear combination of |w| basis functions:**

$$V^{\mathbf{w}}(\mathbf{x}) = \sum_i w_i f_i(\mathbf{x}_i)$$

Hybrid Approximate Linear Programming

- **Hybrid approximate linear programming (HALP) formulation:**

$$\text{minimize}_{\mathbf{w}} \quad \sum_i w_i \alpha_i$$

$$\text{subject to}: \quad \sum_i w_i F_i(\mathbf{x}, \mathbf{a}) - R(\mathbf{x}, \mathbf{a}) \geq 0 \quad \forall \mathbf{x} \in \mathbf{X}, \mathbf{a} \in \mathbf{A}$$

Basis function relevance weight:

$$\alpha_i = \sum_{\mathbf{x}_D} \int_{\mathbf{x}_C} \psi(\mathbf{x}) f_i(\mathbf{x}) d\mathbf{x}_C$$

Difference between the basis function $f_i(\mathbf{x})$ and its discounted backprojection:

$$F_i(\mathbf{x}, \mathbf{a}) = f_i(\mathbf{x}) - \gamma \sum_{\mathbf{x}'_D} \int_{\mathbf{x}'_C} P(\mathbf{x}' \mid \mathbf{x}, \mathbf{a}) f_i(\mathbf{x}') d\mathbf{x}'_C$$

HALP solutions

HALP formulation contains infinite number of constraints,

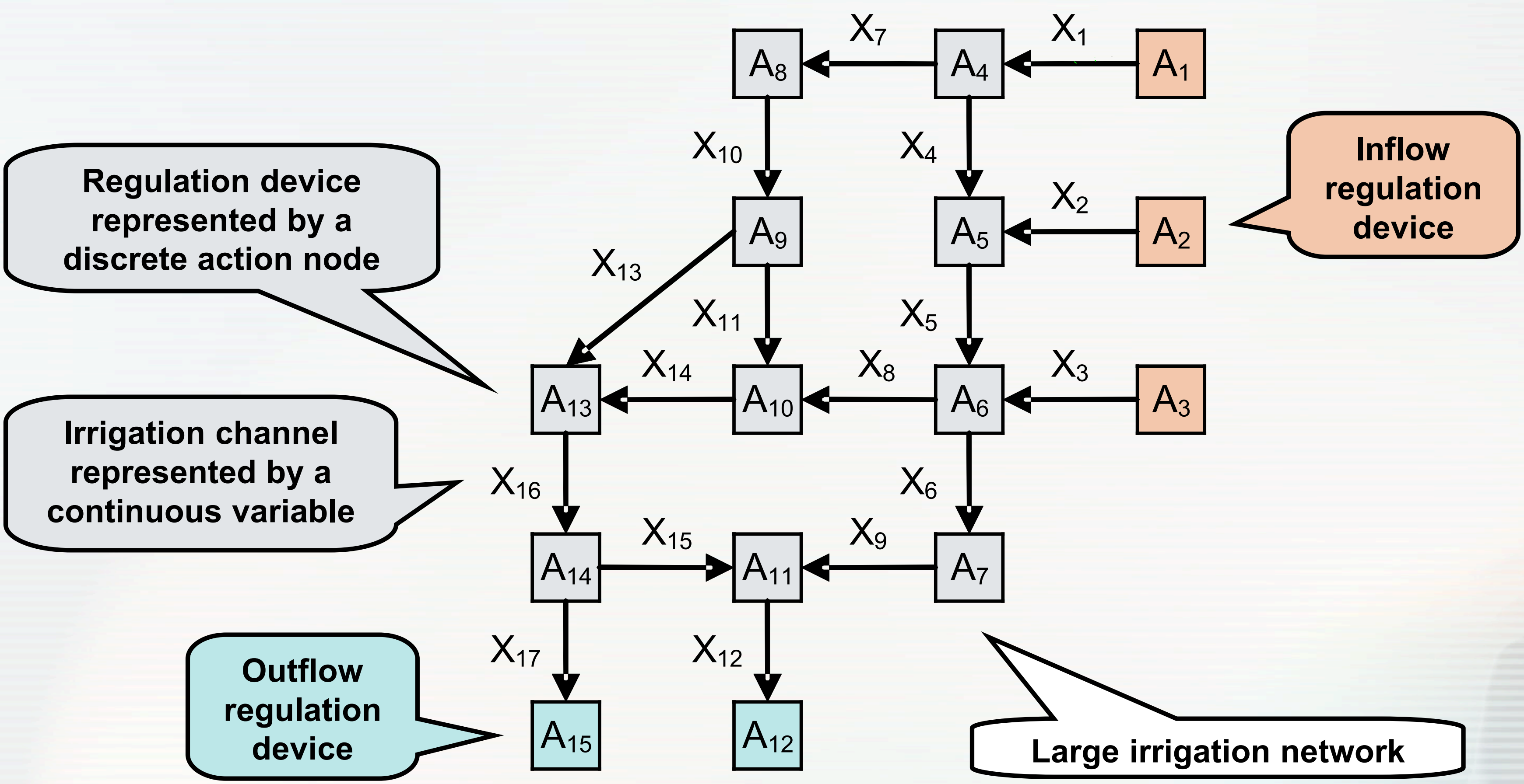
$$\sum_i w_i F_i(\mathbf{x},\mathbf{a}) - R(\mathbf{x},\mathbf{a}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{X}, \mathbf{a} \in \mathbf{A}$$

Methods:

- **Monte Carlo** (Hauskrecht, Kveton, 2004):
  - **Sample constraint space**
- ε**–HALP** (Guestrin, Hauskrecht, Kveton, 2004):
  - **Discretize constraint space on the regular grid**
  - **Take advantage of the structure**
  - **Cutting plane methods on discrete state space**
  - **Exponential in the treewidth of the problem**
- **Direct cutting plane methods** (Kveton, Hauskrecht, 2005)
  - **Stochastic optimizations for finding a maximally violated constraint**

Scale-up potential

- **Problems with over 25 state and 20 action variables**



Traffic Management

- Recent surveys show
  - Americans spend billions of hours in city traffic jams
  - Congestions have grown in urban areas of every size
  - Congestions have increasing economic impact
  - The problem is too complex and growing rapidly
  - No technology has provided a definite solution yet
- Our approach
  - Automatic traffic management to ease congestions
  - Traffic management exhibits local structure with long-term global consequences

Automatic Traffic Management