Probabilistic Robotics

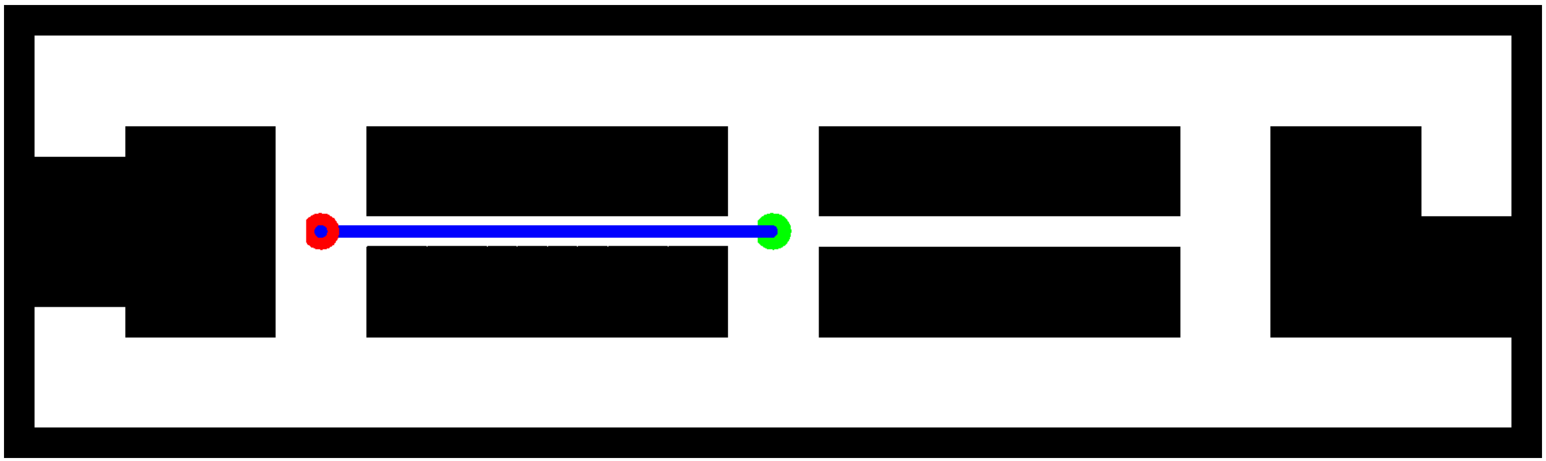
Planning and Control:

Markov Decision Processes

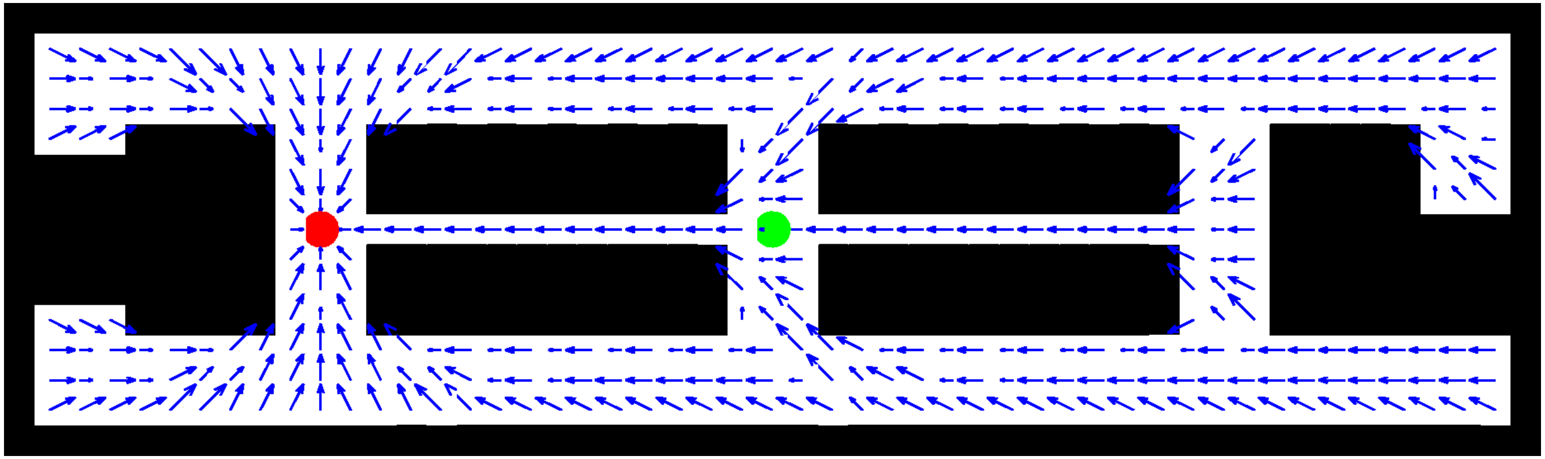
Problem Classes

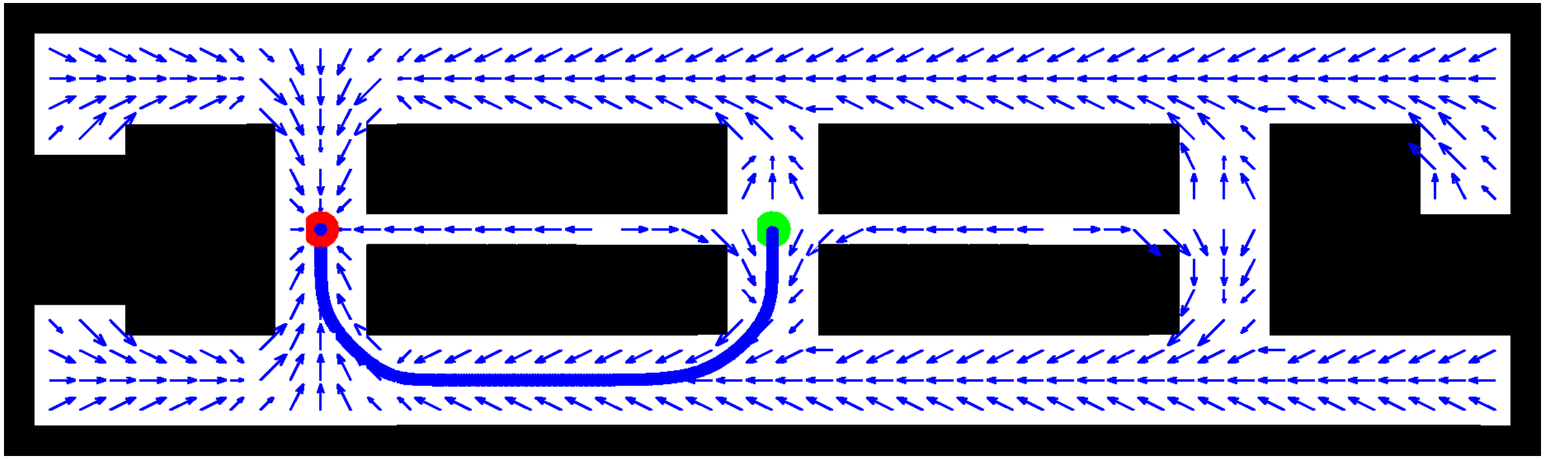
- Deterministic vs. stochastic actions
- Full vs. partial observability

Deterministic, fully observable

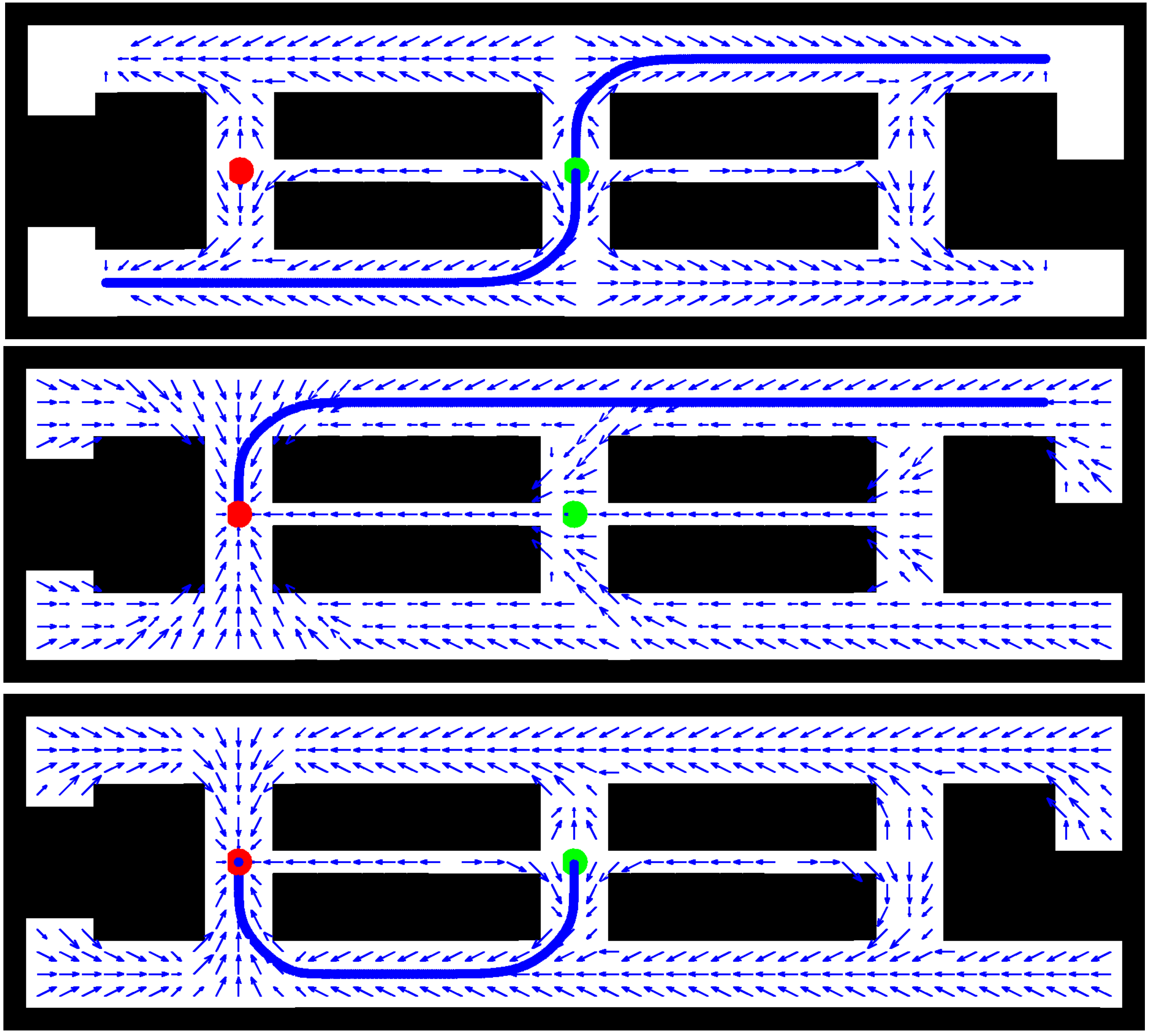


Stochastic, Fully Observable

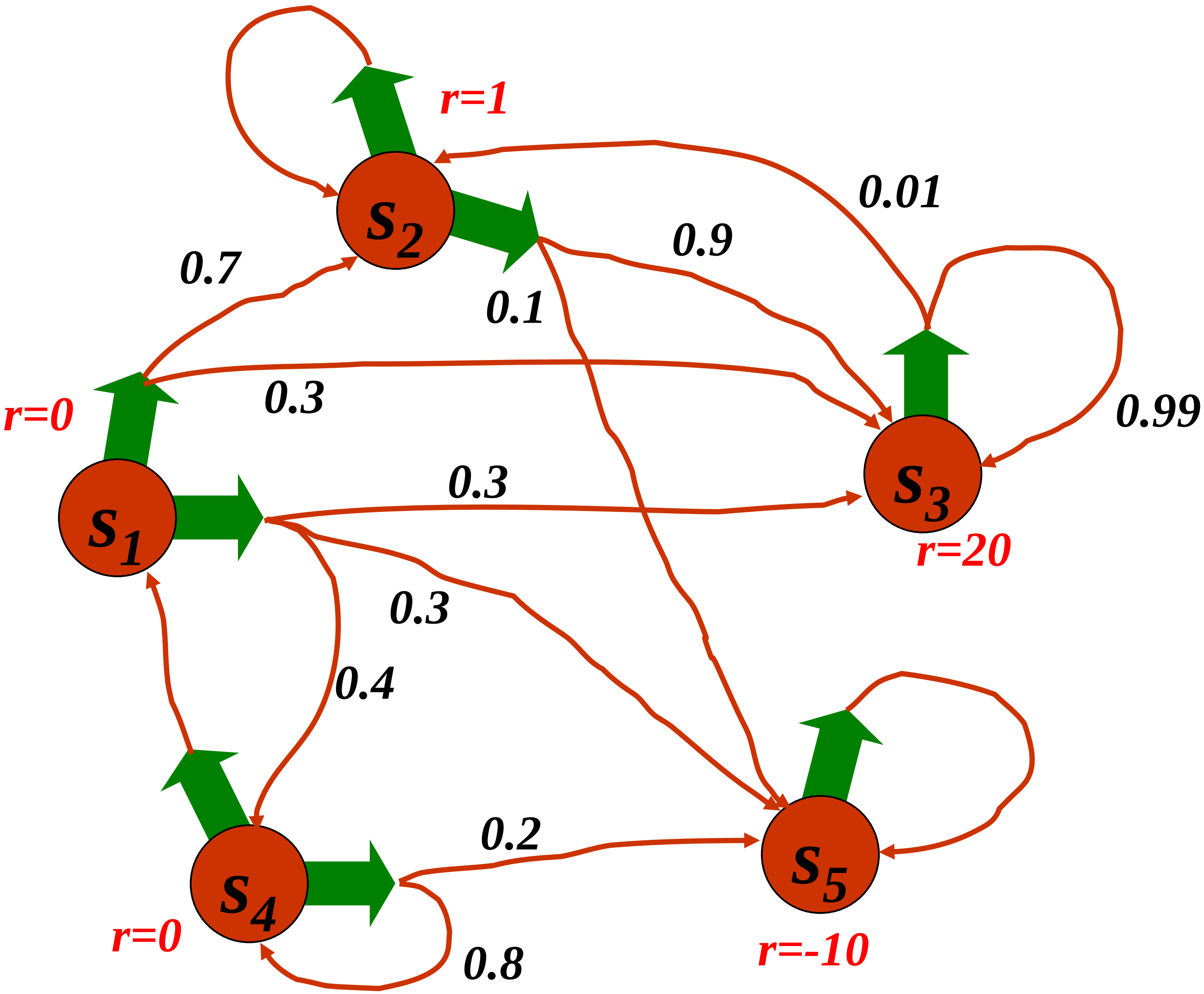




Stochastic, Partially Observable



Markov Decision Process (MDP)



Markov Decision Process (MDP)

- **Given:**
- States x
- Actions u
- Transition probabilities $p(x'|u,x)$
- Reward / payoff function $r(x,u)$

- **Wanted:**
- Policy $\pi(x)$ that maximizes the future expected reward

Rewards and Policies

- Policy (general case):

$$\pi: \quad z_{1:t-1}, u_{1:t-1} \rightarrow \; u_t$$

- Policy (fully observable case):

$$\pi: \quad x_t \rightarrow \; u_t$$

- Expected cumulative payoff:

$$R_T = E \left[\sum_{\tau=1}^{T} \gamma^{\tau} r_{t+\tau} \right]$$

  - T=1: greedy policy
  - T>1: finite horizon case, typically no discount
  - T=infty: infinite-horizon case, finite reward if discount < 1

Policies contd.

- Expected cumulative payoff of policy:

$$R_T^{\pi}(x_t) = E\left[\sum_{\tau=1}^{T} \gamma^{\tau} r_{t+\tau} \mid u_{t+\tau} = \pi\left(z_{1:t+\tau-1} u_{1:t+\tau-1}\right)\right]$$

- Optimal policy:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \quad R_T^{\pi}(x_t)$$

- 1-step optimal policy:

$$\pi_1(x) = \underset{u}{\operatorname{argmax}} \quad r(x,u)$$

- Value function of 1-step optimal policy:

$$V_1(x) = \gamma \max_{u} \; r(x,u)$$

2-step Policies

- Optimal policy:

$$\pi_2(x) = \operatorname*{argmax}_u \left[r(x,u) + \int V_1(x')\, p(x' \mid u, x)\, dx' \right]$$

- Value function:

$$V_2(x) = \gamma \max_u \left[r(x,u) + \int V_1(x')\, p(x' \mid u, x)\, dx' \right]$$

T-step Policies

- Optimal policy:

$$\pi_T(x) = \underset{u}{\operatorname{argmax}} \left[r(x,u) + \int V_{T-1}(x')\, p(x' \mid u, x)\, dx' \right]$$

- Value function:

$$V_T(x) = \gamma \max_u \left[r(x,u) + \int V_{T-1}(x')\, p(x' \mid u, x)\, dx' \right]$$

Infinite Horizon

- Optimal policy:

$$V_\infty(x) = \gamma \max_u \left[r(x,u) + \int V_\infty(x')\, p(x' \mid u, x)\, dx' \right]$$

- Bellman equation
- Fix point is optimal policy
- Necessary and sufficient condition

Value Iteration

- for all *x* do

$$\hat{V}(x) \leftarrow r_{\min}$$

- endfor

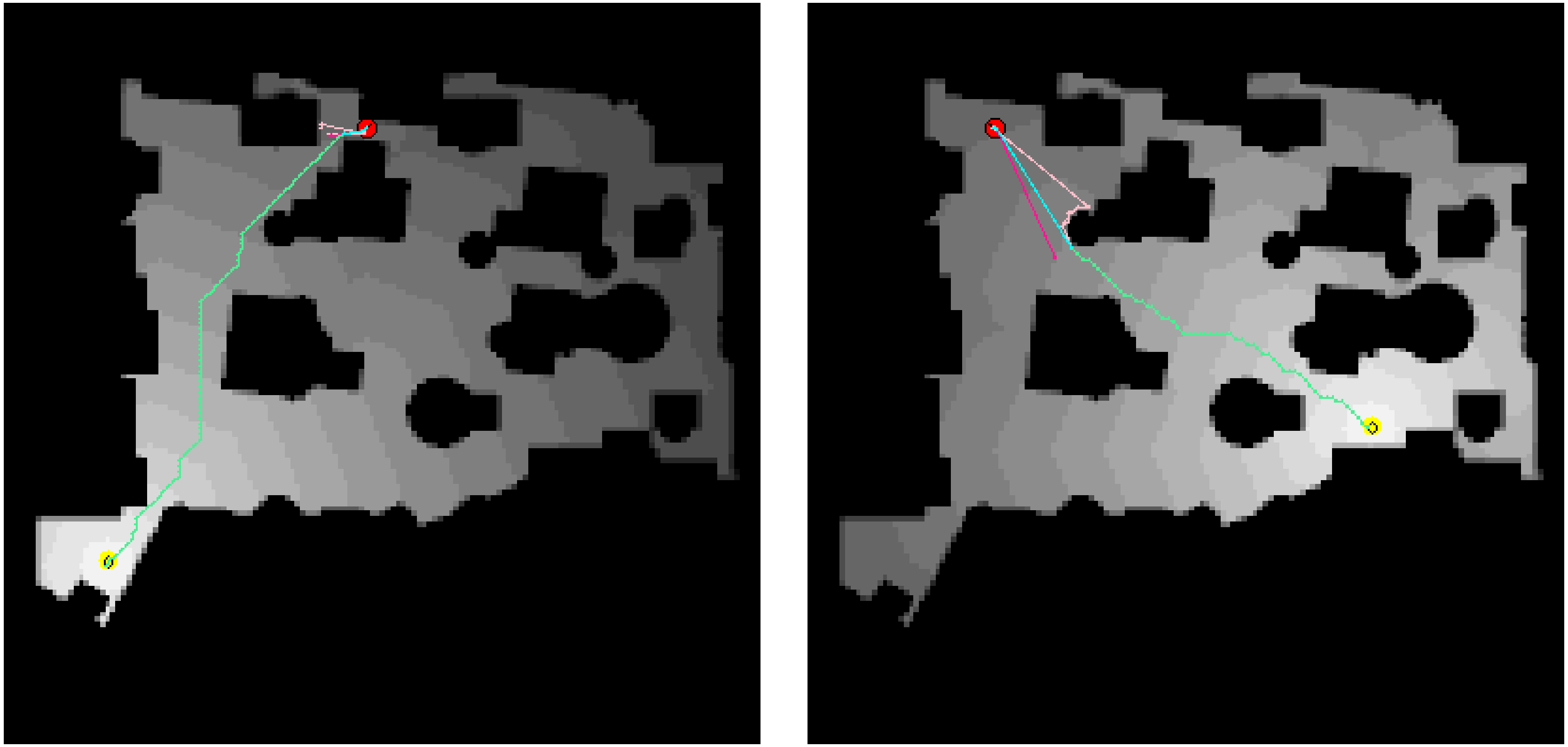
- repeat until convergence
  - for all *x* do

$$\hat{V}(x) \leftarrow \gamma \max_u \left[r(x,u) + \int \hat{V}(x') \, p(x' \mid u, x) \, dx' \right]$$

  - endfor
- endrepeat

$$\pi(x) = \operatorname*{argmax}_u \left[r(x,u) + \int \hat{V}(x') \, p(x' \mid u, x) \, dx' \right]$$

Value Iteration for Motion Planning



Value Function and Policy Iteration

- Often the optimal policy has been reached long before the value function has converged.
- Policy iteration calculates a new policy based on the current value function and then calculates a new value function based on this policy.
- This process often converges faster to the optimal policy.