a_t: action

Actor

observation

action

Episode cost

Lagrange multiplier

Simulation

Goal

Agent

Hazard

Vases

Critic

observation

value

Cost Critic

observation

value