4/21: model-based RL

Announcements:
- HW4 due Thurs, 4/23
- Guest lecture, Prof. Abhishek Cauligi, 4/30

Last time:
- sysID

Today:
- model-based RL
- CBFs

Last time:

$H(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = \tau$

"system identification"

$D = \{(x_t, u_t, x_{t+1})\}$

$\min_\theta \sum_{t=0}^{T-1} \frac{1}{2} \| x_{t+1} - f(x_t, u_t; \theta) \|^2 = \sum_{t=0}^{T-1} r_t(\theta)^2$

↑ myosco ⇒ CEM

θ: when does $\theta^* \to \theta_{true}$

---

e.g.: (robo-chef)

goal: use a fork to pick and twirl spaghetti

u: joint torques

y: encoders, image

x: ? $q, \dot{q}, T_{base}, v_{base}$ T_{fork}, v_{fork}

1. parameters ?
2. state space
3. state estimation

things that are hard to simulate:
- fluids
- flexibles
- people

---

another approach: what if we let our simulator be a neural net?

$\min_\theta \sum_{t=0}^{T-1} \| x_{t+1} - f_\theta(x_t, u_t) \|^2$

↑ deep network

typical name for this: "model-based RL"

Alg:
1. do MPC with current model $\hat{f}_\theta$
2. apply those actions on the real system
3. Add new trajectories $\{(x_t, u_t, x_{t+1})\}$ to dataset; update $\hat{f}$

[Nagabandi '18] Berkeley — [Dreamer '20] Deepmind

Important design decisions:

1. Horizon

suppose f_θ is good for single-step

solutions: $x_{t+1:T} = f_\theta(x_t, u_{t:T-1})$

aside: works for physics-based

$\hat{x}_{t+1} = f_\theta(x_t, u_t)$

$\hat{x}_{t+2} = f_\theta(\hat{x}_{t+1}, u_{t+1})$

$\frac{\partial x_{t+1}}{\partial \theta} = \frac{\partial f_{t+1}}{\partial \theta} + \frac{\partial f_{t+1}}{\partial x_{t+1}} \frac{\partial x_{t+1}}{\partial \theta}$

2. modeling uncertainty

[Nagabandi] $\hat{x}_{t+1} \sim \mathcal{N}(\mu_\theta(x_t, u_t), \Sigma)$ ↑ possibly learned

$\max_\theta \sum_i \log p(x_{t+1} | x_t, u_t, \theta)$

why? 1. often our system is stochastic
2. including ensembles is good for measuring confidence

3. Handling observations

state estimator $y_{1:t}, u_{0:t-1} \to \hat{x}_t$ ⇒ optimal control $\hat{x}_t \to u_{t:T-1}, x^*_{t:T}$

"world model"

$\{(y_0, u_0, y_1, \cdots, y_t)\}$ ← observations, not actions

$z_t = \phi(y_t)$ "image encoder"

$z_{t+1} = f_\theta(z_t, u_t)$ $\min_{\phi,\theta} \| y_{t+1} - \phi^{-1}(f_\theta(\phi(y_t), u_t)) \|$

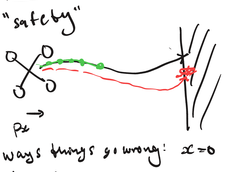
$\hat{y}_{t+1} = \phi^{-1}(z_{t+1})$

---

physics	vs.	world models
+ right structures (conservation of mass/energy)		− don't exhibit object permanence
+ very sample efficient		− sample inefficient
− observations		+ visual observations
− modeling everything		+ generalize to physics that is hard to model

---

"safety"



$h(x) \geq 0$

$p_x \leq 0 - \varepsilon$

↳ put in traj. opt.

ways things go wrong: $x = 0$
1. model error
2. MPC can be greedy

answer: add safety constraints @ multiple levels of stacks