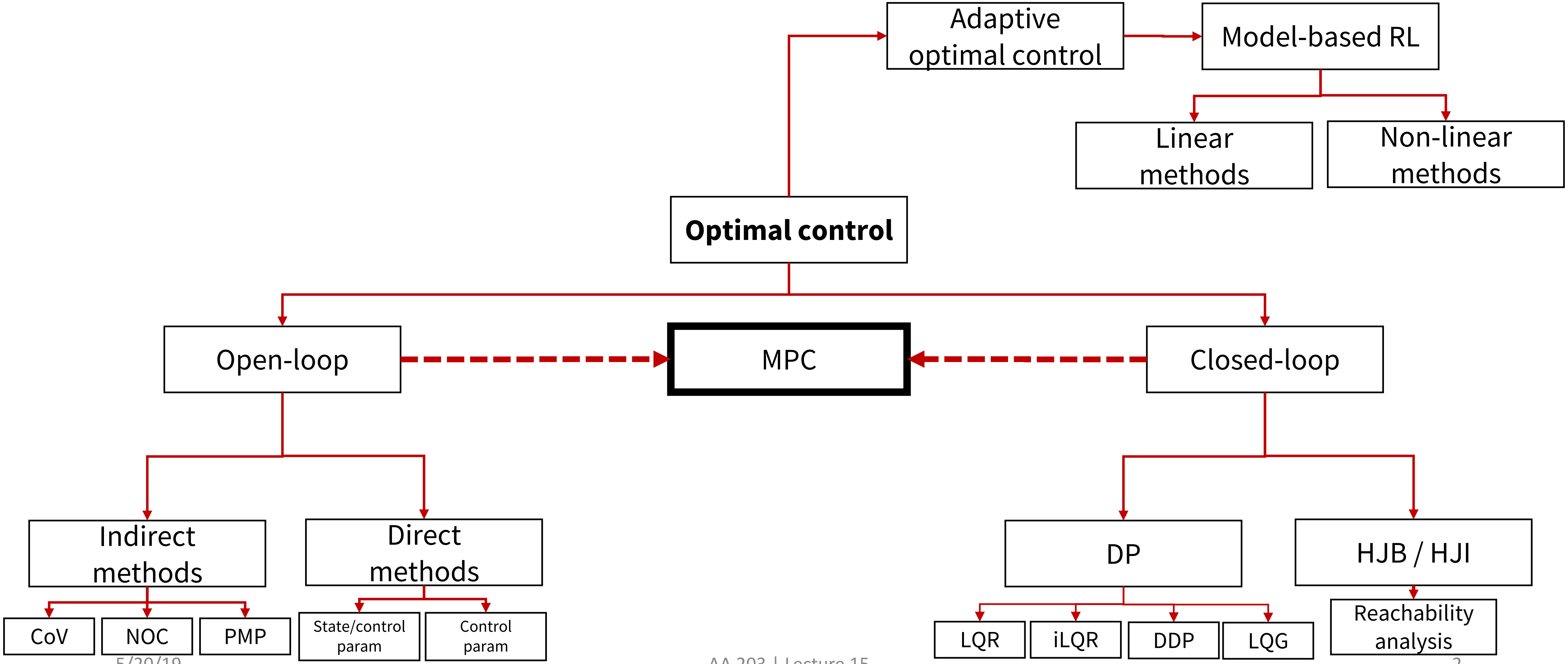
AA203
Optimal and Learning-based Control

Adaptive Optimal Control

Stanford University

Autonomous Systems Lab

Roadmap


Adaptive optimal control
Model-based RL
Linear methods
Non-linear methods
Optimal control
Open-loop
MPC
Closed-loop
Indirect methods
Direct methods
DP
HJB / HJI
CoV
NOC
PMP
State/control param
Control param
LQR
iLQR
DDP
LQG
Reachability analysis

Problem statement

- Up until now, we have aimed to control a (possibly stochastic) system

$$f(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\omega}_k)$$

under a given cost function, subject to state or action constraints.

- For remainder of the class, we will look at controlling a system of the form

$$f(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\omega}_k; \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is an **unknown** vector of parameters governing state evolution

Approaches

If we don't know our exact state evolution, what should we do? Many options:

- In many cases, when the unknown parameters have only a small effect, a feedback controller will adequately compensate for model error
- We can use robust control approaches (e.g. minimax control strategies)
- We can use observed state transitions to attempt to estimate $\boldsymbol{\theta}$ or improve our control strategy

What can we learn?

- We can directly attempt to estimate $\boldsymbol{\theta}$, then use optimal control strategies to plan a controller given the model
  - Commonly referred to as indirect adaptive control or model-based reinforcement learning
- Learning the value function is not useful because it is not actionable, but can learn the Q function

$$Q(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{E}\left[c(\boldsymbol{x}, \boldsymbol{u}) + J(\boldsymbol{x}')\right]$$

  and choose actions via maximizing
- Can directly learn the policy, π
  - Direct adaptive control or model-free reinforcement learning

How does learning happen?

- We will mostly look at the case in which we attempt to learn θ, but will touch briefly on the other cases
- Two possible learning settings: one episode or multiple
  - **One episode**: want to learn and re-optimize our controller online. This is the standard setting for adaptive control
  - **Multiple episodes**: interact with the environment in episodes, in which the system is reset at the start of each episode. Learning and policy optimization can happen between episodes. This is the standard setting for reinforcement learning.
  - **Zero episodes**: the system identification approach, in which learning is done based on data gathered before operation

System identification as LBC

- For many problems, we don’t need to learn online
- A standard control engineering pipeline is to do experiments in advance to build a data-driven model of the dynamics
- Then, we can use this model for planning and control without further learning
- Relies on having an engineer in the loop in learning, designing experiments, resetting the system, etc.

Linear Regression

- Given system of form

$$y = \boldsymbol{\phi}'(\boldsymbol{x})\boldsymbol{\theta} + \epsilon$$

where $\boldsymbol{\theta}$ is a vector of parameters, and ϵ are iid zero-mean constant variance errors.

- Want to find $\hat{\boldsymbol{\theta}}$ to minimize squared error criterion

$$e = \sum_{k=0}^{N} \|\boldsymbol{y} - \Phi\hat{\boldsymbol{\theta}}\|_2^2$$

where $\boldsymbol{y} = [y_0, \ldots, y_N]'$ and $\Phi = [\phi(\boldsymbol{x}_0), \ldots, \phi(\boldsymbol{x}_N)]'$

Linear Regression

- Solution to minimization problem takes form

$$\hat{\boldsymbol{\theta}}^* = (\Phi'\Phi)^{-1}\Phi'\boldsymbol{y}$$

- Gauss-Markov theorem: $\hat{\boldsymbol{\theta}}^*$ is the **best linear unbiased estimator** (for any noise distribution that obeys assumptions)
- If noise distribution is Gaussian, $\hat{\boldsymbol{\theta}}^*$ is the maximum likelihood estimator
- Bayesian perspective: $\hat{\boldsymbol{\theta}}^*$ is *maximum a posteriori* estimator given uninformative prior. For Gaussian noise, we can compute the posterior over $\boldsymbol{\theta}$ in closed form for a Gaussian prior.

System identification via linear regression

- As an example, consider a generic robotic manipulator, which has dynamics

$$M\ddot{\boldsymbol{x}} + C(\boldsymbol{x}, \dot{\boldsymbol{x}})\dot{\boldsymbol{x}} + G(\boldsymbol{x}) = B(\boldsymbol{x})\boldsymbol{u}$$

- We can discretize this via forward Euler to yield

$$M\frac{\dot{\boldsymbol{x}}_{t+1} - \dot{\boldsymbol{x}}_t}{dt} + C(\boldsymbol{x}_t, \dot{\boldsymbol{x}}_t)\dot{\boldsymbol{x}}_t + G(\boldsymbol{x}_t) = B(\boldsymbol{x}_t)\boldsymbol{u}_t$$

- Note that this is **nonlinear in the state**, but **linear in** M, and so we can use least squares to identify the inertial parameters
- Practically, least squares can be written in recursive form for efficiency

Performance questions

The system identification approach leads to several questions:

- How much data is required to learn the model? How can we quantify a “good” estimate? We care about controller performance, not model accuracy, so do we require an accurate model?
- How should we design the inputs used for data collection? What if an engineer can’t intervene to prevent system failure during data collection?
- What if our system does not fall in the class of systems we are considering?

Adaptive control

- Broadly, adaptive control aims to perform online adaptation of the policy to improve performance
- This can be done via directly updating the policy or updating the model and re-optimizing or re-computing the controller
- Most adaptive control work does not consider the *optimal adaptive control* problem; they focus on proving stability of the coupled controller and adaptive component

Adaptive control approaches

Encompasses a huge variety of techniques:

- Adaptive pole placement or policy adaptation (direct adaptive control)
- Iterative learning control
- Gain scheduling
- Model reference adaptive control (MRAC)
- Model identification adaptive control (MIAC)
- Dual control

Model identification adaptive control

- We will focus on the latter two: MIAC and dual control
- MIAC simply combines model estimation with a controller that uses the estimated model
- Important distinction between *certainty-equivalent* and *cautious* approaches
  - **Certainty-equivalent:** maintains point estimate of model and uses that model for policy selection/optimization. Note that unlike the LQG setting, certainty-equivalence is sub-optimal.
  - **Cautious:** Maintains measure of estimator uncertainty, incorporates the uncertainty into the controller. This is often overly robust because it does not account for future info gain!

Dual control

- Most adaptive control is “passive”: it does not incorporate the value of information or actively explore
- Dual control augments the state with the estimate of the unknown parameters, and uses the joint dynamics
- By performing DP in this “hyperstate”, can find a controller that optimally probes/explores the system; optimally trades off exploration/exploitation
- Practically, designing dual controllers is difficult, so sub-optimal exploration heuristics are used: more on this later!

Example: Adaptive control of the LQ system

- Consider simplest possible linear-quadratic discrete-time system, of the form

$$x_{k+1} = ax_k + bu_k + \epsilon$$

with cost $c(x, u) = qx^2 + ru^2$

- One of the first proposed approaches was Simon [1956], in the context of inventory control
- He proposed a certainty-equivalent approach, where a, b are recursively computed via least squares, and the optimal policy is computed with respect to those estimates

Example: Adaptive control of the LQ system

- Becker et al. [1985] showed that this approach can converge to to incorrect estimates with positive probability, which can lead to suboptimal performance!
- Alternative approach: dual control. We will assume that the only unknown parameter is b to further simplify the problem. Then, the hyperstate is $\boldsymbol{z}_k = [x_k, \hat{b}_k]'$. Note that the problem is immediately nonlinear, so can not use LQR for control design.
- Bar-Shalom and Tse [1976] solve the problem for two timesteps, but a general solution is difficult. This is the simplest possible problem!
- Indeed, certainty equivalence holds if and only if there is no dual effect [Bar-Shalom and Tse, 1973]!

Example: Adaptive control of the LQ system

- Passive adaptive control may converge to a poor solution and dual control is largely intractable, so common practice is to turn to exploration heuristics
- To achieve convergence of the estimator, we require $(\Phi'\Phi)^{-1} \to 0$ as $t \to \infty$ [Lai and Wei, 1982].
- Many approaches to guaranteeing asymptotic convergence

Example: Adaptive control of the LQ system

- A variety of schemes exist for performing exploratory actions
  - White noise injection into control [Becker et al., 1985]
  - “Strategic” injection of white noise into control [Lai and Wei, 1987]: inject white noise into control at certain times
  - Cost-biased estimator (as opposed to maximum likelihood) [Campi and Kumar, 1998]: adds term to estimator that biases model estimate toward lower cost models; effectively “optimistic”
  - “Bet on the best”: act wrt best non-falsified model (extreme optimism) [Bittanti and Campi, 2006]
- These methods all analyzed in terms of asymptotic convergence. Recently, considerable work on finite-horizon performance, see [Abbasi-Yadkori and Szepesvari, 2011] or recent work by Recht et al.

Next time

- Reinforcement learning for linear systems