Goal: using toy system, do a “comprehensive computational study illustrating different tradeoffs”

Along that vein, we’ve set up a simulation environment for a two-state system

$$x_{t+1} = Ax_t + Bu_t + w$$

$$A = \begin{bmatrix} 1+\alpha & 0 \\ 0 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$w \sim \mathcal{N}(0,\sigma)$$

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ for all trajectories}$$

Each trajectory takes exactly 5 steps

Observations from basic RL experiments:
-More uncontrollability impairs ability to converge (1)
-More instability takes longer to converge (unclear whether ultimate performance changes, see 2)
-More noise decreases end performance (3)

Todo:
1. Theoretical result on convergence? (not for control audience)
2. Run longer experiments
3. Normalize all performance metrics by control LQR optimal performance***, given explicit instantiation
    1. ***Working out the details of how to evaluate performance, RL is finite horizon evaluation but we care about infinite horizon for theoretical comparisons

Moving to bigger picture:

We’re seeing a story come out of the comparisons. From model-based to model-free, we want to compare performance for the following:

1. Totally model-based, everything pre-known (including noise, plant, etc) LQR
2. Naive sysID on plant (and noise?)+assume correct ID & do ^^LQR
3. Nik’ stuff: do control that’s robust to specific errors in ID
4. Model-ish-based RL: output a linear controller (easy to evaluate quality in theoretical terms)*~*
5. Model-Free RL: Policy gradient methods

~ use control metrics as RL reward function?

This seems like a nice Control-RL tutorial in a classroom setting

For ACC:
Either: 1 long paper, need better metrics to visualize “convergence”, “bias”, and “variance” of each experiment on a single plot
Or: Sister papers #1,2 #4,5 (3?)