Last Time

- ☐ HJ Reachability
- ☐ code demo!
- ☐ robust safety

Lecture 5

EAIS SP'25

Andrea Bajcsy

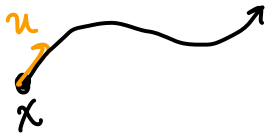
This Time:

- ☐ dynamic games
- ☐ code demo!
- ☐ (reinforcement learning approx.)

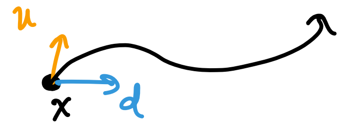
Announcement: HW #1 due Wed!

Robustifying Safety

BEFORE: → NOW:



$\dot{x} = f(x, u)$



$\dot{x} = f(x, u, d)$

"ego" "robot" "disturbance"

ROBUST BACKWARDS REACHABLE TUBE (BRT) of a set $\mathcal{F} \subset \mathcal{X}$ and dynamical system $\dot{x} = f(x, u, d)$ is:

$$BRT(t) := \{ x \in \mathcal{X} : \underbrace{\forall u(\cdot) \in \mathbb{U}_t^T}_{\text{for all things the robot could do}}, \underbrace{\exists\, d(\cdot) \in \mathbb{D}_t^T}_{\text{there is something the disturbance could do ...}}, \underbrace{\mathbf{x}_{x,t}^{u,d}(\tau) \in \mathcal{F}}_{\text{that leads the robot to failure}} \text{ for some time } \tau \in [t, T] \}$$

This is the set of all starting states from which no matter the controller's effort, the disturbance can push system into $\mathcal{F}$.

The way we will formulate an "optimal control" problem whose solution represents this unsafe set will be via:

$$\underbrace{\text{zero-sum}}_{\text{there is a winner + loser}} \quad \underbrace{\text{dynamic}}_{\text{game evolves over time}} \quad \underbrace{\text{games}}_{\text{result/outcome depends on 2+ players/inputs}}$$

Our "game" formulation could look something like this:

$$V(x,t) \equiv \max_{u(\cdot)\in \mathbb{U}_t^T} \; \min_{d(\cdot)\in \mathbb{D}_t^T} J(x, u(\cdot), d(\cdot))$$

$$\text{s.t. } \dot{x}(\tau) = f(x(\tau), u(\tau), d(\tau)) \quad \forall \tau \in [t, T]$$

$$u(\tau) \in \mathcal{U}$$

$$d(\tau) \in \mathcal{D}$$

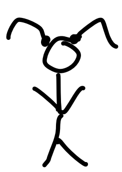
Here, the robot (u) is trying to maximize the objective $J(\cdot,\cdot,\cdot)$ and the other player (d) is minimizing. In other words, the robot is optimizing the worst-case objective

Importance of Information Patterns

When we have 2 players reacting to each other, their optimal strategy will depend on what information they each have access to.

$\boxed{\text{Example}}$ Suppose there are 2 boxes, each with 2 slots. each slot contains prize money. Player A (You ü) wants to maximize prize money while Player B (competition organizer) wants to minimize player A's prize money.

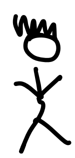
Box 1		Box 2	
2,000	10	1	6,000
s1	s2	s1	s2



Player A

$u \in [\text{Box } 1, \text{Box } 2]$

you choose Box

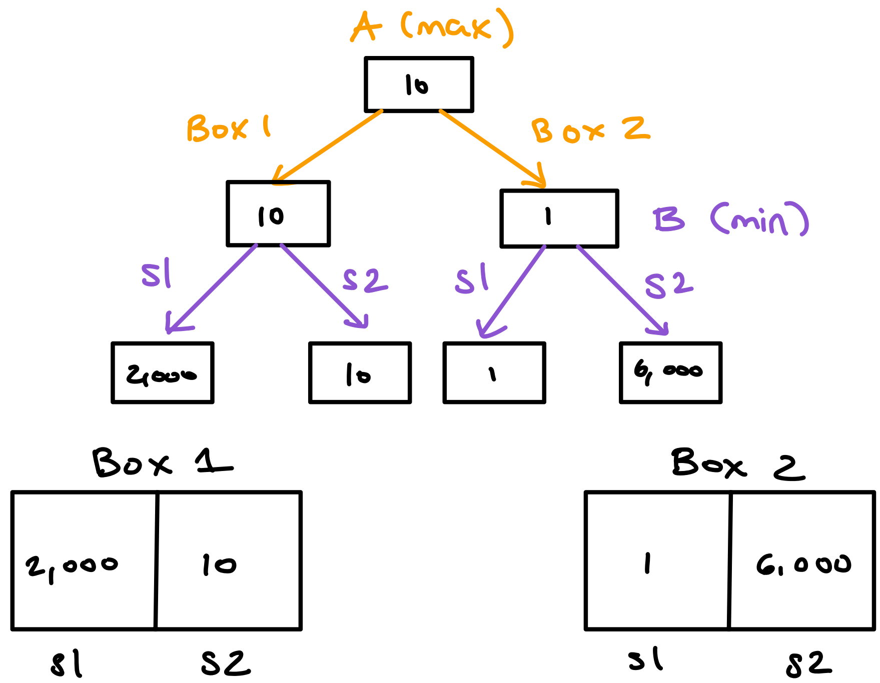


Player B

$d \in [\text{slot } 1, \text{slot } 2]$

they choose slot

Suppose player A goes first:



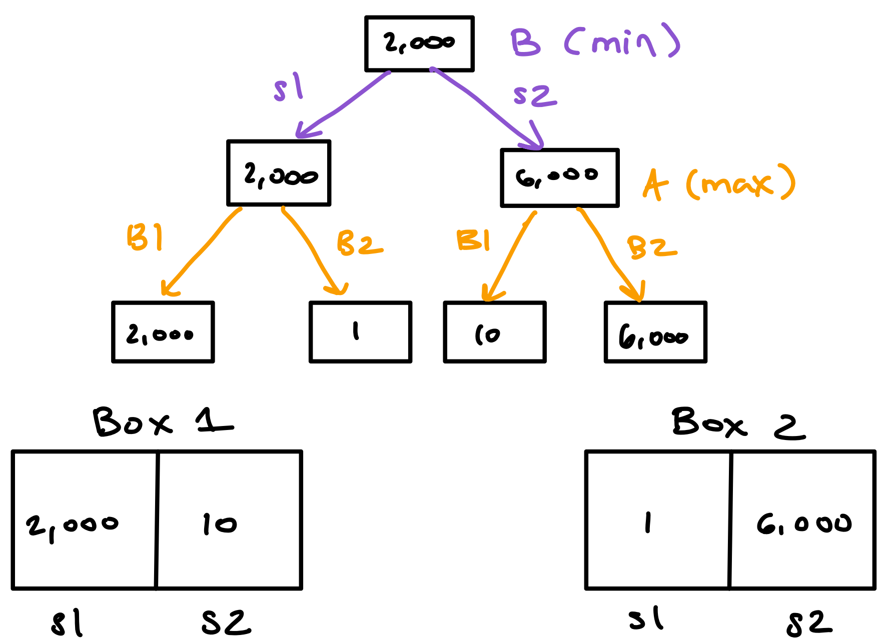
Best outcome for A is to pick Box 1 & get Rew = $10.

Mathematically:

$$\underbrace{\max_{u\in[B1,B2]}}_{\text{"computed 2nd" but "plays first"}} \Big(\underbrace{\min_{d\in[s1,s2]} J(u,d)}_{\text{"computed first" but "plays 2nd"}} \Big) = 10$$

⇒ "best response" strategy

Suppose Player B goes first:



Best strategy for player B is to pick slot 1 and pay player A a reward of \$2,000.

$$\min_{d \in [S1, S2]} \left(\max_{u \in [B1, B2]} J(u, d) \right) = \$2{,}000$$

This phenomenon we just saw can be stated via

minimax inequality:

$$\max_{a} \left(\min_{b} J(a,b) \right) \leq \min_{b} \left(\max_{a} J(a,b) \right)$$

Von Neumann 1928: equal when A, B are compact, convex sets + $J(\cdot, b)$ is concave for fixed b

$J(a, \cdot)$ is convex for fixed a

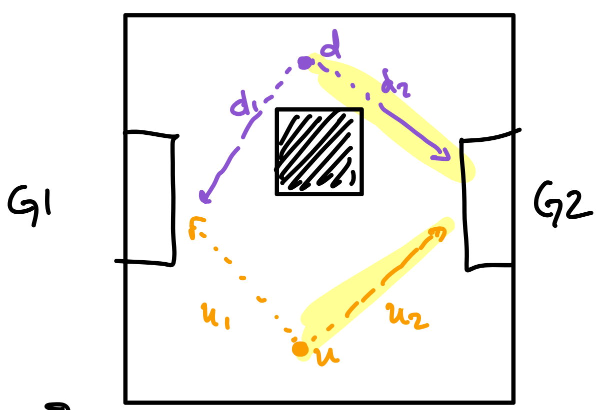
In dynamic games the outcome depends on WHEN and WITH WHAT INFORMATION each player decides their inputs

NOW.

just talked about this (order of play)

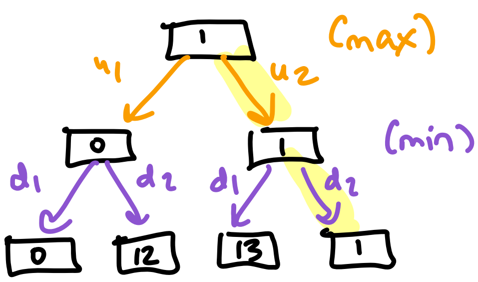
OPEN-LOOP INFO PATTERN:

$$\max_{u(\cdot) \in \mathbb{U}_t^T} \min_{d(\cdot) \in \mathbb{D}_t^T} J(x, u(\cdot), d(\cdot))$$



Here u wants to reach G1 OR G2 without being intercepted by d.

BUT d wants to intercept u.



} time $[0, T]$

} time $[0, T]$

J ↪ "dist btwn. agents"

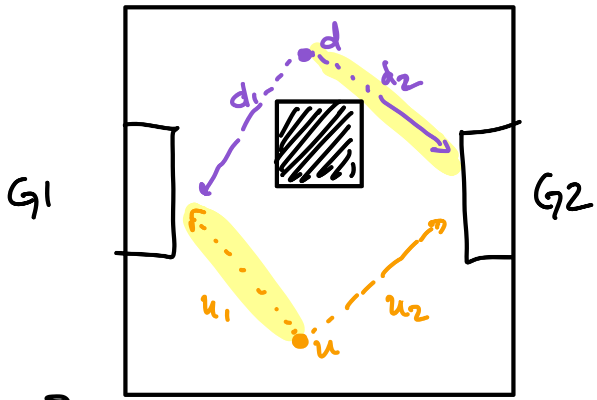
$u(\cdot)$ "declares" entire signal and then $d(\cdot)$ gets to see this and intercepts!

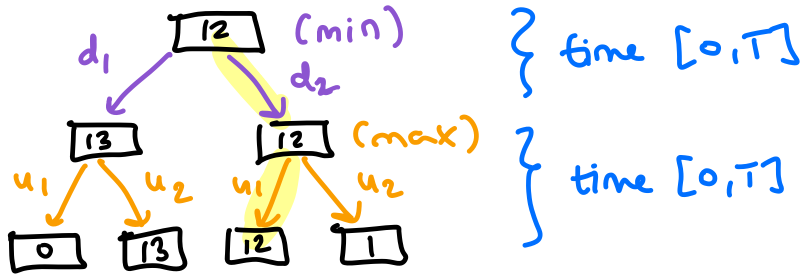
(*) OVERLY PESSIMISTIC!

let's swap the order of play again, but we are playing over continuous signals (or sequences of decisions):

OPEN-LOOP INFO PATTERN:

$$\min_{d(\cdot) \in \mathbb{D}_t^T} \; \max_{u(\cdot) \in \mathbb{U}_t^T} J(x, u(\cdot), d(\cdot))$$





J ↺ "dist btwn. agents"

$d(\cdot)$ "declares" entire signal and then $u(\cdot)$ gets to see this and picks the other goal!

⊛ OVERLY OPTIMISTIC!

CLOSED-LOOP INFORMATION PATTERNS

The above formulation is not suitable for many practical systems. We would like the controller (u) to ADAPT over time as system evolves, BUT respect the fact that @ any time t, we only have information up to time t.

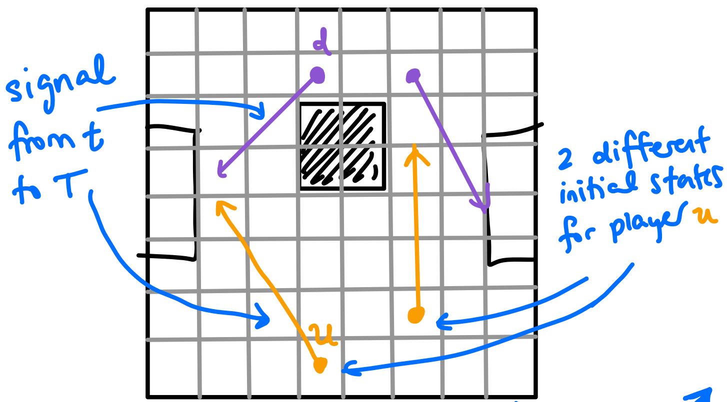
We can model this by solving for feedback policies:

$$V(x,t) \equiv \max_{\pi_u} \min_{\pi_d} J(x, \pi_u(\cdot), \pi_d(\cdot))$$

$\pi_u : X \to U$

$\pi_d : X \to D$

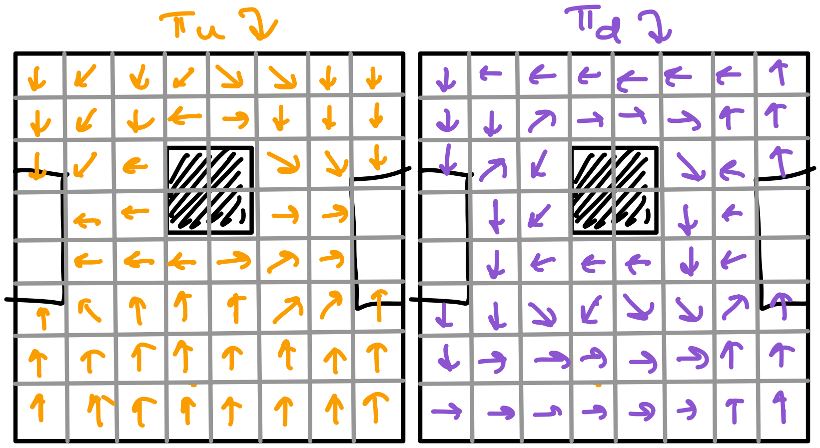
Before $\left(\max_{u(\cdot)\in\mathbb{U}_t^T} \ \min_{d(\cdot)\in\mathbb{D}_t^T} \cdots\right)$



from any initial state (denoted by • icon), each agent has to commit to entire control signal or sequence that they cannot change in future.

$\Rightarrow$ this is $\boxed{\text{open-loop}}$

Now $\left(\max_{\pi_u} \ \min_{\pi_d} \cdots\right)$



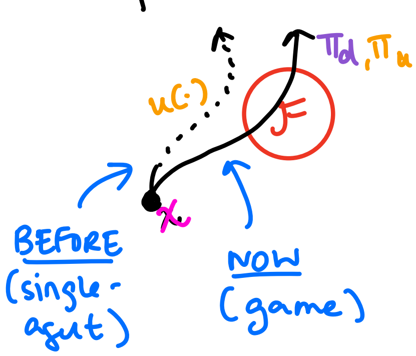
for each initial state, the players commit to a policy — or a map from each state to their action.

$\Rightarrow$ this is $\boxed{\text{closed-loop}}$

⊛ NOTE: for this illustrative example, since u and d are 2 separate agents, the state space is $\mathcal{X} := \mathcal{X}_u \times \mathcal{X}_d$ and policies map from where BOTH players are, to actions:

$$\pi_u(x_u, x_d) \to u \quad , \quad \pi_d(x_u, x_d) \to d$$

Bringing this back to safety analysis, we just have to chose a objective function which lets us remember the closest we ever got to failure:



$$J(x, \pi_u, \pi_d, t) := \underbrace{\min_{\tau\in[t,T]} \ell\left(\mathbf{x}_{x,t}^{\pi_u \pi_d}(\tau)\right)}_{\text{minimum future distance influenced by } u \text{ and } d}$$

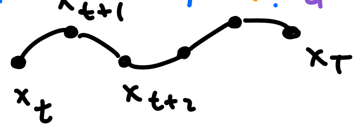
We can apply the game-theoretic principle of optimality called the <u>Tenet of Transition</u> by Rufus Isaacs:

"If play proceeds from one position (state) to a second, and V is thought of as known to the second, then it is determined at the first by demanding that players optimize (i.e. make minimax) the increment of V during the transition."

$\boxed{\text{example}}$ (discrete-time)

$$V_t(x_t) := \min_{\pi_u(x)} \max_{\pi_d(x)} \min_{\tau \in \{t, \dots T\}} \ell\left(x_\tau^{u,d}\right)$$

trajectory: w/ π_u & π_d



$$= \min_{u^t} \max_{d^t} \dots \min_{u^T} \max_{d^T} \min_{\tau \in \{t, \dots T\}} \ell\left(x_\tau^{u,d}\right)$$

$$= \min_{u^t} \max_{d^t} \min \Big\{ \ell(x_t), \underbrace{\min_{u^{t+1}} \max_{d^{t+1}} \dots \min_{u^T} \max_{d^T} \min_{s \in \{t+1, \dots T\}} \ell\left(x_s^{u,d}\right)}_{:= V_{t+1}(f(x_t, u_t, d_t)) \text{ by } \underline{\text{Tenet of Transition}}} \Big\}$$

$$= \min_{u^t} \max_{d^t} \min \Big\{ \ell(x_t), V_{t+1}\big(f(x_t, u_t, d_t)\big) \Big\}$$

$$= \min \Big\{ \ell(x_t), \min_{u^t} \max_{d^t} V_{t+1}\big(f(x_t, u_t, d_t)\big) \Big\}$$

Applying this to the continuous or discrete-time zero sum safety critical games we formulated above, we get two key equations that allow us to extend dynamic programming tools to the robust (dynamic game) setting:

b/c it's a game!

Hamilton-Jacobi-$\boxed{\text{Isaacs}}$ Variational Inequality (HJ-VI)

$$\min\left\{\ell(x)-V(x,t),\ \frac{\partial V}{\partial t}+\max_{u\in\mathcal{U}}\boxed{\min_{d\in\mathcal{D}}}\frac{\partial V}{\partial x}\cdot f(x,u,d)\right\}=0$$

$$V(x,T)=\ell(x)$$

add this & solve "subgame" over just instantaneous actions

ROBUST SAFETY BACKUP (discrete-time):

$$V_t(x_t)=\min\left\{\ell(x_t),\ \max_{u_t}\ \min_{d_t} V_{t+1}(f(x_t,u_t,d_t))\right\}$$

$$V_T(x_T)=\ell(x)$$