AA/EE/ME 548: Linear Multivariable Control

Lecture 06

4/16/2025

Sequential Decision Making

Control/decision-making over a horizon: choosing a sequence of actions where each one may have enduring consequences

Variations: deterministic vs. stochastic, full vs. partial observability, discrete vs. continuous, open-loop vs. closed-loop control

What should I spend my time on now so I can have a successful career in the future?

“Life is a POMDP”

Freshman

I have a midterm tomorrow morning. Should I eat a healthy meal, go to sleep early, or study through the night?

I have a homework due in 2 weeks, should I start now or go skiing with friends?

Should I attend that networking event or go home and just Netflix and chill?

Sequential Decision Making – Examples

- **Robotics & Control:** trajectory planning, autonomous vehicle routing, process control/regulation
- **Games:** Go, Chess, Diplomacy, Starcraft
- **Resource allocation:** adjusting investments, inventory as the markets/demand/supply changes
- **Healthcare:** what drugs/treatment to use given potential risks/benefits and patient conditions evolving over time

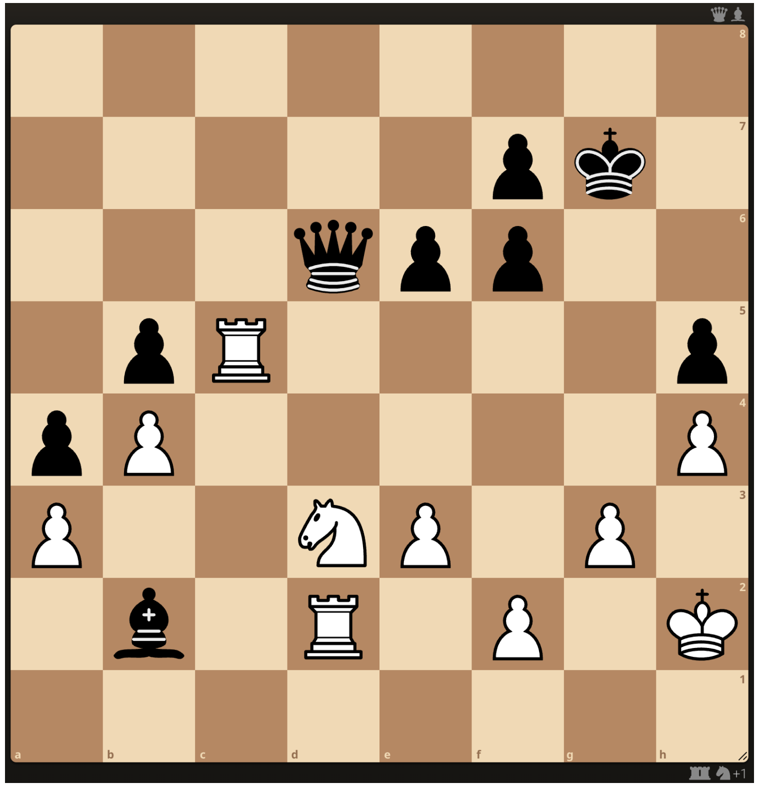
Planning: how do you reason about myriad possible futures to decide what action to take right now?

Motivating Example – Chess

- White to move, Carlsen vs. Nepomniachtchi WCC 2021 Game 6

Who is winning?

How sharp is the position (how critical is White's choice)?



What move should White play?

Motivating Example – Chess

Who is winning?

White is winning. Why?

Motivating Example – Chess

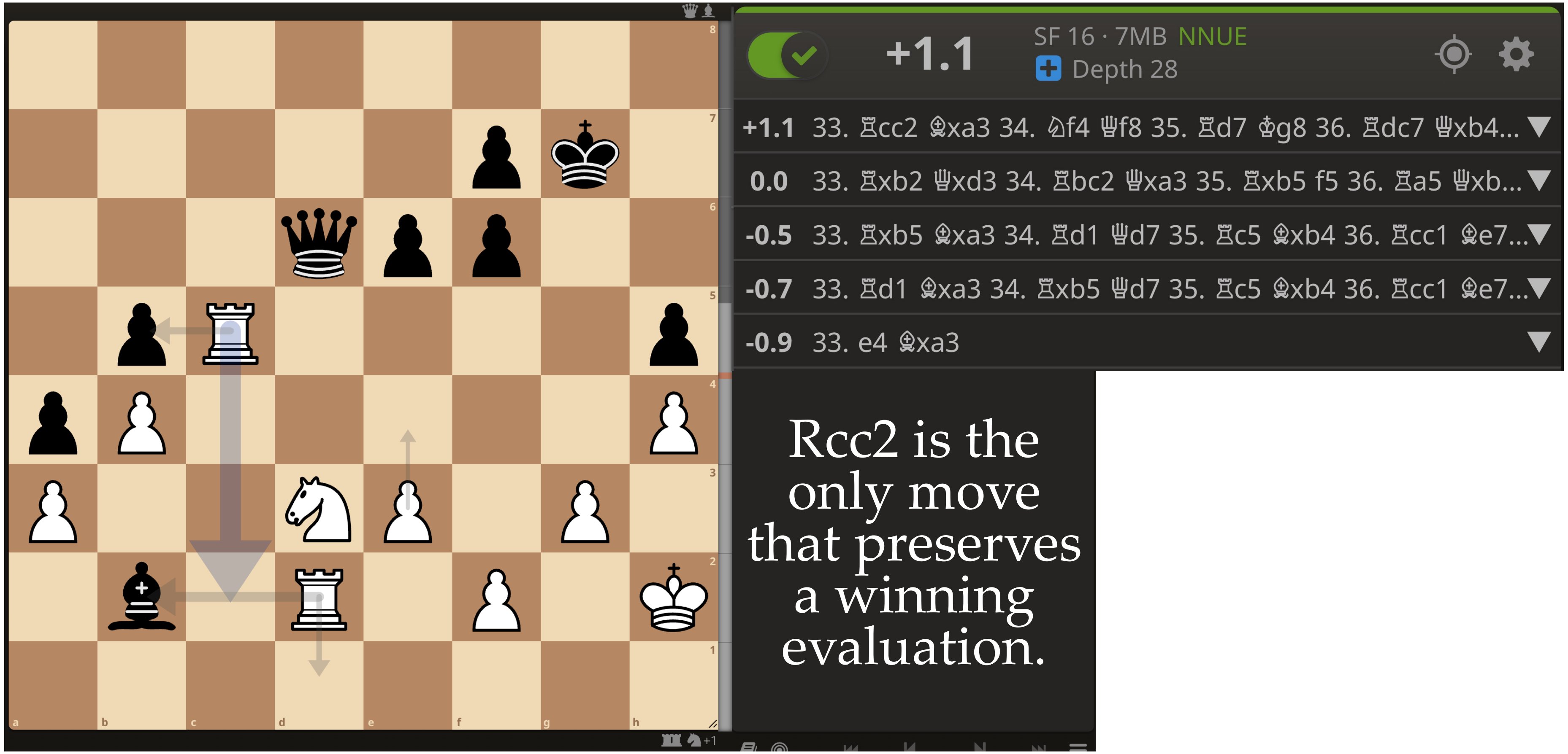
What move should White play?

Rcc2.

Why?

Motivating Example – Chess

How sharp is the position (how critical is White's choice)?



Rcc2 is the only move that preserves a winning evaluation.

Motivating Example – Chess

- Caveats
  - With infinite compute (i.e., infinite "depth")/optimal play, the only possible evaluations are win (∞)/loss ($-\infty$)/draw (0) – think Tic-Tac-Toe
  - There's a second player ("robust control" – minmax analysis, or "stochastic control" – modeling player action distributions)
  - Discrete state space, discrete action space, unusual definition of terminal set/reward a bit different from typical controls applications

Motivating Example – Chess

- Takeaways
  - Decision-making can be approached through the lens of **value functions**
    - Amortizes considerations about myriad possible futures into a single value that informs present decision making
  - (state) Value function "V"
    - From this state, what is the optimal reward (or reward corresponding to following a given policy)?
  - State-action value function "Q"
    - From this state *and taking this action*, what is the optimal reward (or reward corresponding to following a given policy)?

Problem Formulation – Ingredients

- Model
  - State: x_k (or s_k)
  - Action: u_k (or a_k)
  - Dynamics: $x_{k+1} = f(x_k, u_k, k)$

- Objective
  - Cost/reward: $J(x_0) = h_N(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k, k)$

- Policy/controller
  - Open-loop action sequence: $\{u_0, u_1, \dots, u_{N-1}\}$
  - Closed-loop policy: $u_k = \pi(x_k, k)$

How do we select a control policy that optimizes the objective, subject to the system dynamics?

Problem Formulation – Ingredients

- Model
  - State: x_k (or s_k), continuous time $x(t)$, state constraints $x_k \in X_k$
  - Action: u_k (or a_k), continuous time $\mathrm{u}(t)$, control constraints $u_k \in U(x_k, k)$
  - Dynamics: $x_{k+1} = f(x_k, u_k, k)$, continuous time $\dot{x}(t) = f(x(t), u(t), t)$
    - Stochastic transition distribution $p(x_{k+1}|x_k, u_k)$ (or disturbance $x_{k+1} = f(x_k, u_k, d_k, k), d_k \sim p(\cdot\,|x_k, u_k)$)
- Objective
  - Cost/reward: $J(x_0) = h_N(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k, k)$, continuous time $= h(x(T), T) + \int_0^T g(x(t), u(t), t)dt$, infinite horizon $J(x_0) = \sum_{k\geq 0} \gamma^k g(x_k, u_k, k)$
- Policy/controller
  - Open-loop action sequence: $\{u_0, u_1, \dots, u_{N-1}\}$, $\mathrm{u}(t)$
  - Closed-loop policy: $u_k = \pi(x_k, k)$, $\mathrm{u}(t) = \pi(x(t), t)$

How do we select a control policy that optimizes the objective, subject to the system dynamics?

Optimal Control

- Hamilton-Jacobi-Bellman equation
  - Dynamic programming in continuous time
- Reachability analysis
  - Optimal safe sets and corresponding policies
  - Compared to CBFs: optimality at the expense of running full DP through state and time (instead of just verifying a CBF condition point-wise)
- Linear Quadratic Regulator (LQR)
  - Provides an optimal policy, not just an optimal plan!

Today’s Outline

- Optimal substructure & the principle of optimality
- Dynamic programming
  - Algorithm for computing an optimal policy and value functions
- Problem settings
  - Deterministic dynamics
  - Stochastic dynamics: Markov Decision Processes (MDPs)
    - Example: Optimal Stopping
  - Partially-observed MDPs (POMDPs)
- Autonomous vehicles, project ideas, general discussion