Active Observing in Continuous-time Control


Samuel Holt


Alihan Hüyük


Mihaela van der Schaar

NeurIPS 2023
Presentation

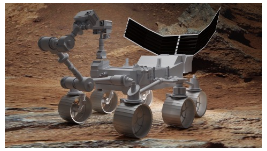
van_der_Schaar
\ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Applications of continuous-time control with observation costs

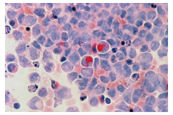
- **Medical cancer chemotherapy treatment**
  - Taking expensive Computed Tomography scans, whilst continuously controlling chemotherapy dosing



- **Mobile robotics**
  - Measuring the robots position, whilst continuously controlling the robot

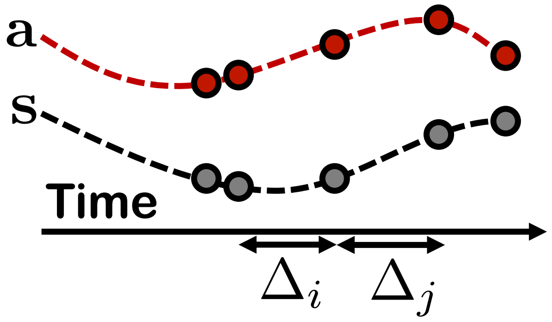
- **Low power communication**
  - Measuring the maximum bandwidth, whilst continuously controlling the channel transmission
- **Biological fish population management**
  - Fish population survey, whilst continuously controlling the food and temperature.

What is continuous-time control with costly observations?

- **Continuous-time environments.** 
  - **Environment dynamics can be described by a *differential equation*** 

$$\frac{\mathrm{d}s(t)}{\mathrm{d}t} = f(s(t), a(t))$$

$$z(t) = s(t) + \varepsilon(t) \qquad \varepsilon(t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$$



  - **Offline dataset trajectories of s and a can be observed at irregular time intervals $\Delta_i \neq \Delta_j$**
- **Observation costs of c** 
- **Objective to maximize utility**

$$\mathcal{U} = \underbrace{\int_0^T r(s(t), a(t), t)\mathrm{d}t}_{\text{Reward } \mathcal{R}} - \underbrace{c|\{t_i : t_i \in [0, T]\}|}_{\text{Cost } \mathcal{C}}$$

What is continuous-time control with costly observations?

Policies ρ, π **interacting with the environment**

1. $t_1 = 0, h_1 = \{(t_1, z(t_1), a(t_1))\}$
2. **For** $i \in \{1, 2, \ldots\}$:
3. **Schedule next observation:** $t_{i+1} = t_i + \rho(h_i)$
4. **Execute actions:** $a(t) = \pi(h_i, t - t_i)$ **for** $t \in [t_i, t_{i+1})$
5. **Take an observation:** $h_{i+1} = h_i \cup \{(t_{i+1}, z(t_{i+1}), a(t_{i+1}))\}$

$$\rho^*, \pi^* = \operatorname{argmax}_{\rho,\pi} \mathbb{E}[\mathcal{U}]$$

First to formalize the problem of continuous-time control whilst deciding when to take costly observations

- **Theoretically, we show that regular observing in continuous time with costly observations is not optimal for some systems and that irregularly observing can achieve a higher expected utility.**

- **Proposition: For some systems, it is not optimal to observe regularly—that is** $\exists f, \sigma_\epsilon, r, c, h, h' : \rho^*(h) \neq \rho^*(h')$

Active Observing Control

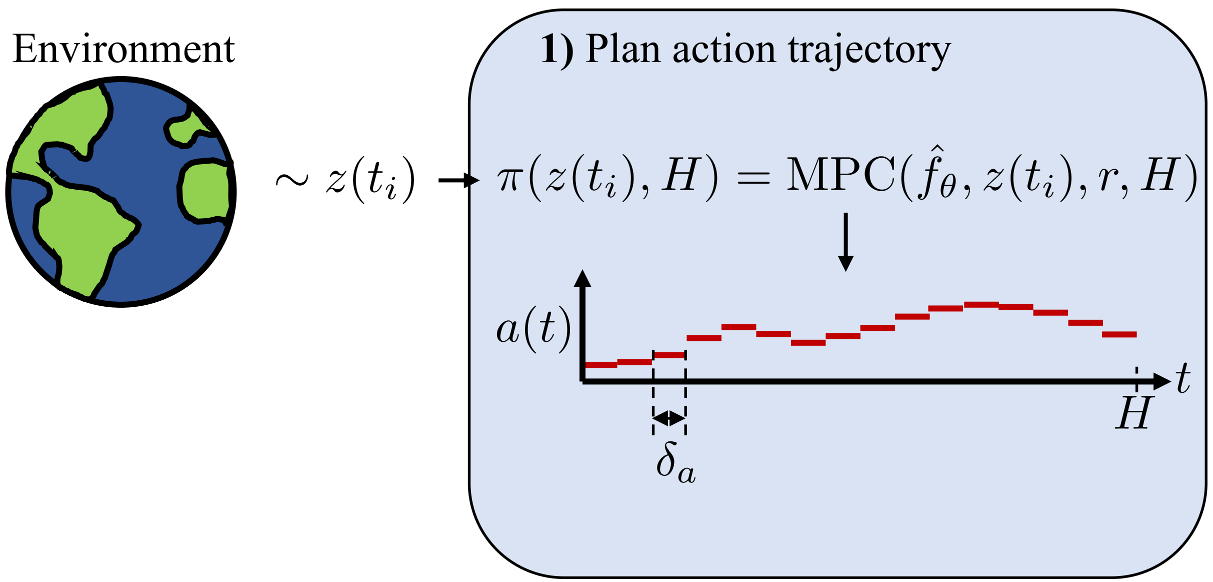
- We propose the **Active Observing Control.**
  - A continuous-time model-based offline RL method that uses a heuristic threshold on the variance of reward rollouts in an model predictive control (MPC) planner.
- **Benefits:**
  - Can avoid discretization errors in time.
  - Can learn from an offline dataset sampled with irregular time intervals and has observation costs.
  - Can achieve high performing utility compared to existing methods.
  - Allows to only observe when it is informative to do so and observe irregularly in time.
  - Is robust to the heuristic threshold hyperparameter.
  - Small run-time complexity, so practical to use.

Active Observing Control

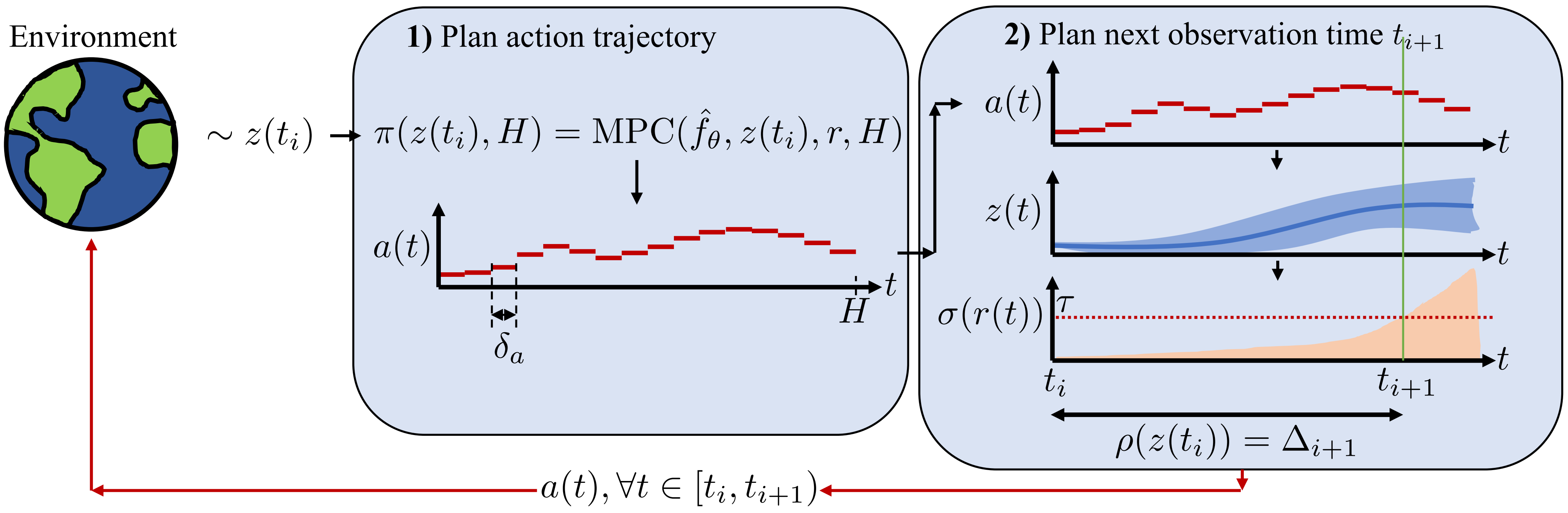
Environment

$\sim z(t_i)$

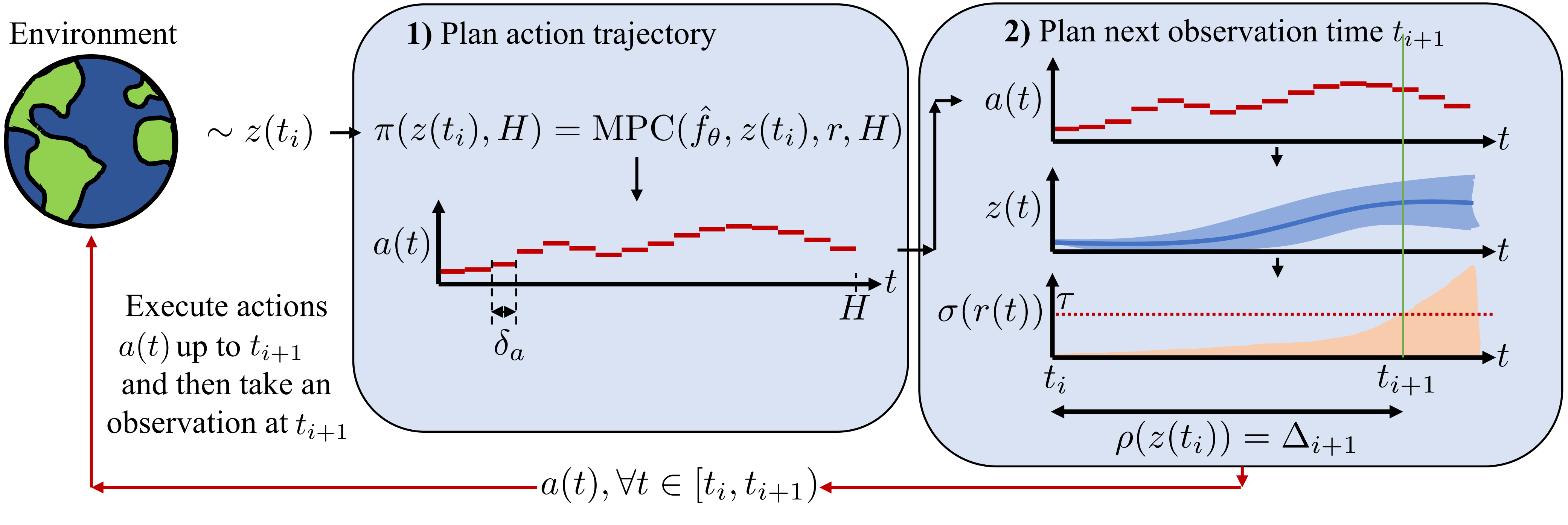
Active Observing Control



Active Observing Control



Active Observing Control



$$\rho(z(t_i)) = \max\{\Delta' \in \mathbb{R}_+ : \sqrt{\mathbb{V}_{z_p}[r(t_i + \Delta')]} < \tau\}$$

Results

- **Normalized utilities $\mathcal{U}$, normalized rewards $\mathcal{R}$ and observations $\mathcal{O}$ of the baselines.**

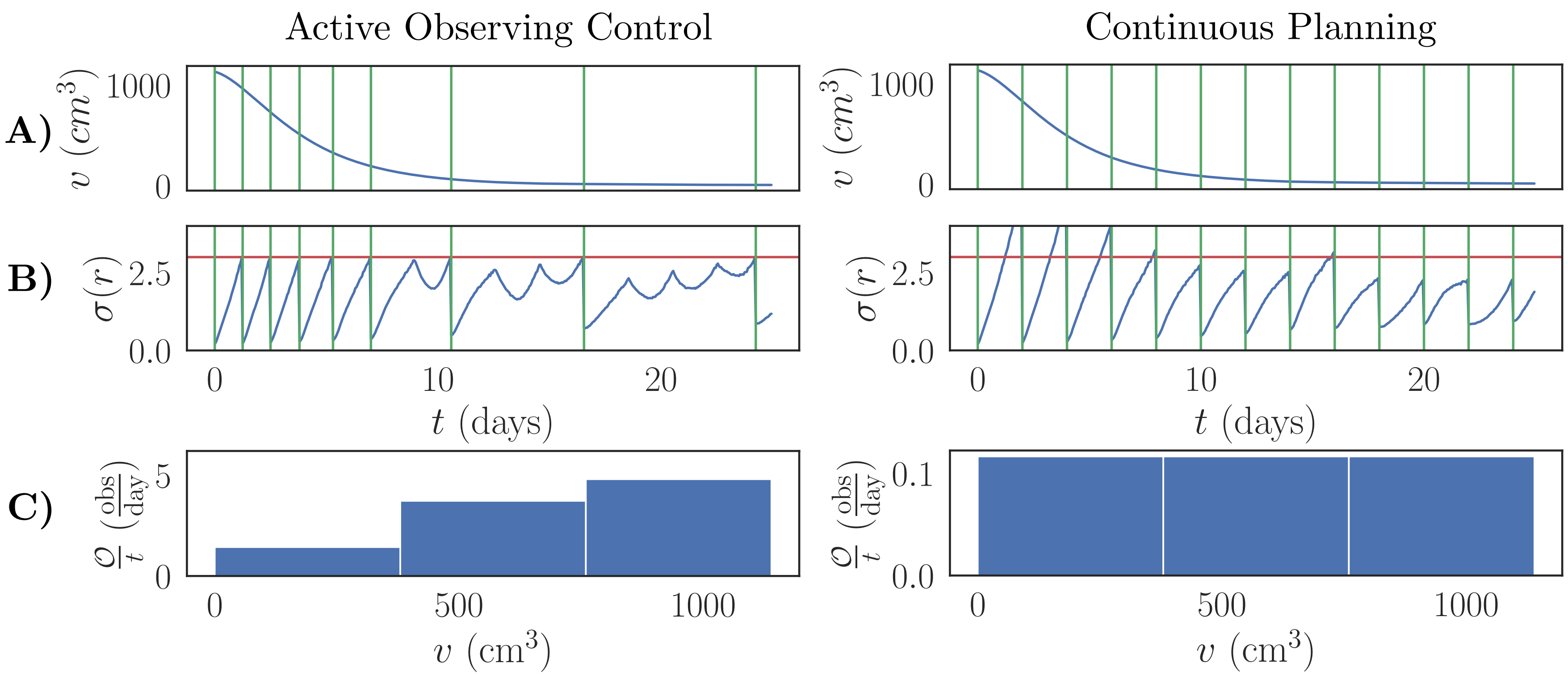
Policy	Cancer			Acrobot			Cartpole			Pendulum		
	$\mathcal{U}$	$\mathcal{R}$	$\mathcal{O}$	$\mathcal{U}$	$\mathcal{R}$	$\mathcal{O}$	$\mathcal{U}$	$\mathcal{R}$	$\mathcal{O}$	$\mathcal{U}$	$\mathcal{R}$	$\mathcal{O}$
Random	0±0	0±0	13±0	0±0	0±0	50±0	0±0	0±0	50±0	0±0	0±0	50±0
Discrete Planning	91.7±0.368	91.7±0.368	13±0	87.1±1.05	87.1±1.05	50±0	83.6±0.56	83.6±0.56	50±0	87.2±0.962	87.2±0.962	50±0
Discrete Monitoring	91±0.532	85.8±0.522	5.08±0.0327	89.6±1.02	80.2±1.14	43.7±0.189	127±0.846	82.9±0.532	42.3±0.107	130±2.52	87.3±0.957	42.1±0.293
Continuous Planning	100±0.153	100±0.153	13±0	100±0.462	100±0.462	50±0	100±0.772	100±0.772	50±0	100±0.904	100±0.904	50±0
Active Observing Control	**105±0.18**	98.8±0.169	3.37±0.0302	**107±0.911**	90.8±0.878	39±0.177	**151±1.54**	99.5±0.774	41.1±0.196	**177±2.18**	98.8±0.912	35.6±0.239

- **Active Observing Control achieves state-of-the-art episodic utility performance across the cancer environment and standard continuous-time RL environments.**
  - **Achieving near expert policy performance, when taking significantly less observations.**

Insight Experiments

How does irregularly observing achieve a higher expected utility than regularly observing?

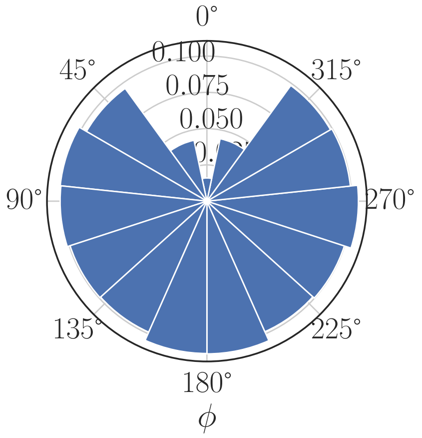
- **AOC Automatically determines to observe larger cancer volumes more frequently as they are more informative, as the future state change magnitude is larger.**



Insight Experiments

How does irregularly observing achieve a higher expected utility than regularly observing?

- **Frequency of observations per state region for Pendulum.**

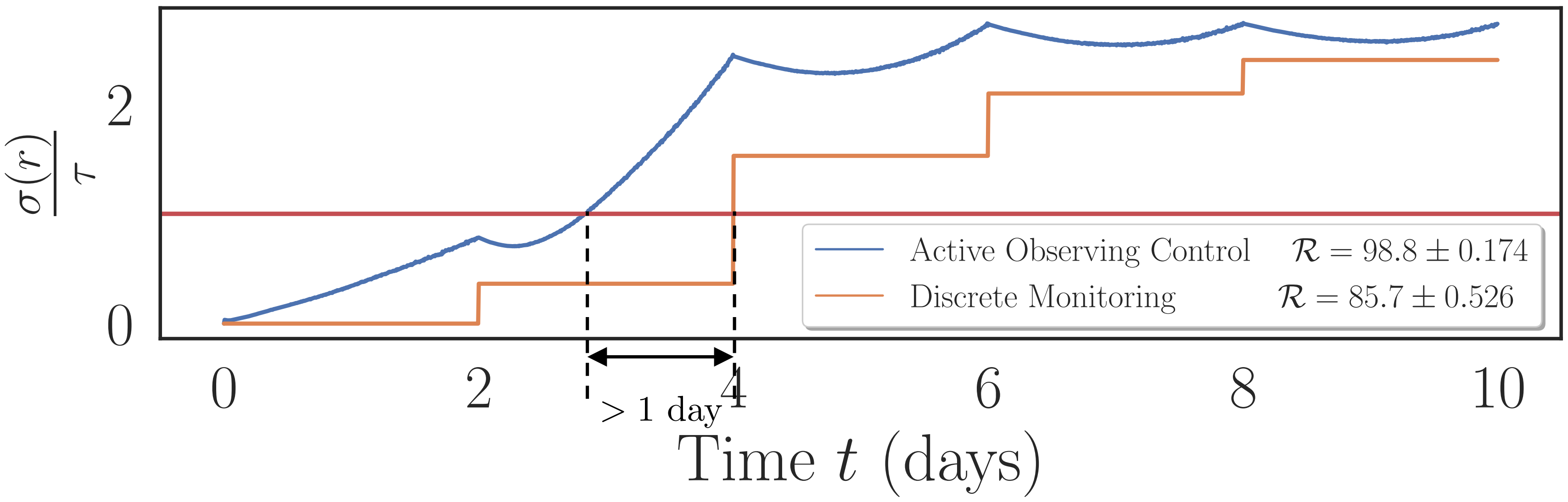


- **Even when Continuous Planning takes the same number of observations as determined by AOC, it still performs worse, because those observations are not *well located*.**

Policy	Cancer $\mathcal{U}$	Cancer $\mathcal{R}$	Cancer $\mathcal{O}$
Active Observing Control	105±0.183	98.8±0.173	3.39±0.0306
Continuous Planning with $\mathcal{O}=3$	102±0.234	95.6±0.234	3±0
Continuous Planning with $\mathcal{O}=4$	103±0.226	97.3±0.226	4±0

Insight Experiments

Why is it crucial to actively observe with continuous-time methods, rather than discrete-time methods?



Contributions

- We are the first to formalize the problem of continuous-time control whilst deciding when to take costly observations.
- Can achieve state-of-the-art utility performance.
- Can correctly observe the state when it is informative to do so.
- It can avoid discretization errors in time, and is robust to its threshold hyperparameter.

- This now **enables**:
  - Dynamic expensive medical scan scheduling
  - New improved methods to build on and solve this real-world applicable costly observing whilst continually controlling problem.

More info

paper

code

neurips.cc/virtual/2023/poster/70479

github.com/samholt/ActiveObservingInContinuous-timeControl

AstraZeneca

Office of Naval Research Science & Technology

van_der_Schaar \ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

sih31@cam.ac.uk

github.com/samholt

samholt.github.io/

linkedin.com/in/samuel-holt