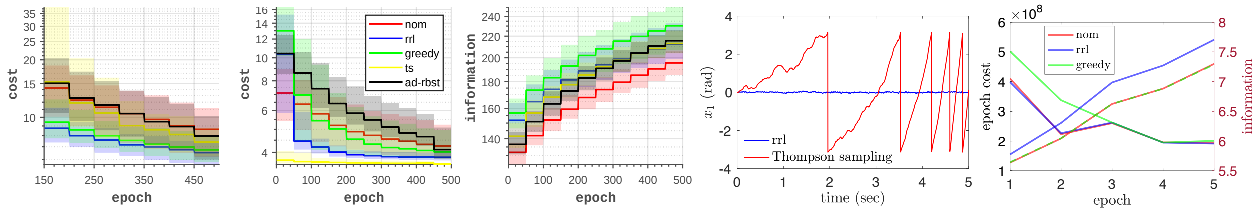
Foremost, we would like to thank the reviewers and (S)ACs for giving up their time to conduct and organize the reviews of our manuscript. Indeed, all reviewers have read the paper thoroughly, and we are grateful for their positive comments and constructive feedback. Moreover, all three reviewers were extremely consistent in their suggestions for improvement: i) include comparisons with SoTA methods, and ii) consider additional experiments.

Comparison with SoTA We now compare the proposed approach with two additional SoTA methods: i) `ad-rbst`: the robust, adaptive method from *Dean et al. Regret Bounds for Robust Adaptive Control of the Linear Quadratic Regulator, NeurIPS 2018*, and ii) `ts`: Thompson sampling. Results are presented in Fig a. Full details will be provided in the revised manuscript (V2), but for now: our proposed method, `rrl`, performs better than the other robust method `ad-rbst`. `rrl` also performs best in terms of worst-case cost (which it was designed to optimize), although `ts` achieves the lowest cost on the true system, **when the policy is stable**. Note that the worst-cost for `ts` is infinite (unstable) in 99%, 54%, and 8% of trials at epochs one, two and three respectively. We also applied Thompson sampling to the hardware-in-the-loop (HIL) experiment, and observed that it was unable to stabilize the system, cf. Fig b.



(a) Simulated example: worst-case (L), true system (C), information (R). (b) Instability in HIL. (c) Simulated pendulum.

Experiments Please allow us to first justify the use of the HIL experiment. The motivation for use of the servomechanism was twofold: i) control of servomechanisms is a ubiquitous task in practice (e.g. robotics), ii) more importantly, unlike e.g. the inverted pendulum, which is nonlinear, the planar servo can be reasonably well modeled (globally) as a linear system. We believe that the servo represents a good compromise between the complexities of a real world system (subtle nonlinearities such as backlash and friction, unmodeled dynamics, disturbances, etc) and a system that approximately satisfies the assumptions of the method. Furthermore, with $n_x = 7$ & $n_u = 2$, the HIL system is not of trivial dimension; many real-world systems can be represented with models of this size. Nevertheless, we also applied the method to a (simulated) inverted pendulum (modeling the Quanser pendulum); cf. Fig c. Although reasonable performance is observed, the results are not as ‘clean’ or ‘explainable’ due to the more complex behavior of the nonlinear system (which does not satisfy the assumptions of the method). At the recommendation of the reviewers, we can include these results in the paper (including experiments on the real physical pendulum), however, we believe the HIL experiment is more compelling, as it is more consistent with the assumptions of the method.

Reviewer #1 We wish to thank R#1 for identifying a number of areas in which the clarity of the manuscript could be improved. All of the following points will be clarified in the revised manuscript (V2). ***Re: measurability of the states,*** as in a number of recent papers (e.g. [11,12,13,16]), we assume that the states are directly measurable. Output feedback control is indeed an important and common scenario, but this is beyond the scope of the paper. In the example, we assume that the filter provides a suitable approximation of the velocity, and that any discrepancies can be partially accounted for by the process noise. ***Re: static-gain policies,*** such policies are popular in practice, due to the simplicity of synthesis and implementation; for instance, the (ubiquitous) proportional-derivative (PD) controller can be implemented with static-gain, and by introducing an ‘artificial state’ (with known dynamics) to the system, integral action (for PID) can be incorporated too. ***Re: comparisons to DP,*** we will add a brief discussion on the challenges of ‘gridding’ continuous state/action spaces in order to apply DP-based methods, citing relevant literature. As suggested, we will report computation times in V2. ***Re: the approximations in §4.1,*** we attempted to discuss each approximation in lines 150-156 and 159-160. However, as you rightly point out, given the importance of such assumptions, they will be discussed in greater detail in V2. ***Re: the outliers in Fig 2a,*** This is an interesting question. Note that while all methods were ‘robust’, sometimes the worst-case cost is quite bad (depending on the data realization). ***Re: Fig 2c and interpolation,*** it is true that greedy first applies the ‘nominal’ controller; however, it also adds exploration ‘noise’ such that greedy and RRL have the same worst-case cost at the first epoch. We must apologize for a small error: the center figure in 2c is actually the cost on the true system. This is why the cost of greedy and RRL differ at the first epoch.

Reviewer #2 Thank you for your positive comments and useful suggestions; please cf. ‘§SoTA’ and ‘§Experiments’.

Reviewer #3 We wish to thank R#3 for the many useful suggestions to improve the manuscript. ***Re: the conclusion,*** we apologize; this was simply due to space restrictions. We will endeavor to include a conclusion in V2. ***Re: the use of the word ‘essential’,*** we completely agree; we will amend this (to ‘certain physical systems’). ***Re: the uniform prior,*** this is an excellent suggestion; we will explain that the prior is degenerate. ***Re: the hardware experiment,*** the synthetic system was introduced to increase the dimensionality of the system so as to make the set-up somewhat more interesting/realistic. ***Re: robustness of the HIL experiment,*** please cf. Fig b for consequences of instability. ***Re: further numerical experiments,*** for real-hardware experiments, please see above. Concerning more complicated simulations, e.g. navigating obstacles, this is a very interesting suggestion. We feel that this goes beyond the scope of the present paper, as it would necessitate higher-level path planning, but is an interesting direction for future work.