A Study of Regret Minimization for Static Scalar Nonlinear Systems

Ying Wang, KTH Royal Institute of Technology
Division of Decision and Control Systems
Advisor: Håkan Hjalmarsson, Collaborators: Mirko Pasquini, Kévin Colin

Motivation & Research Goals

- We study exploration in a static scalar nonlinear optimization problem with an unknown parameter learned from noisy data.
- The goal is to balance exploration and exploitation via regret minimization over a finite horizon.
- The theoretical results suggest that the optimal strategy is either:
  - Lazy exploration: no exploration;
  - Immediate exploration: exploration only at the first time instant.
- A quadratic numerical example illustrates these findings.

Problem

Setting: Static unconstrained scalar optimization problems

$$u_0^* = \arg\min_{u\in\mathbb{R}} \Phi(u,\theta_0)$$
$$\text{s.t. } y_t = h(u_t,\theta_0) + e_t, e_t \sim N(0,1)$$

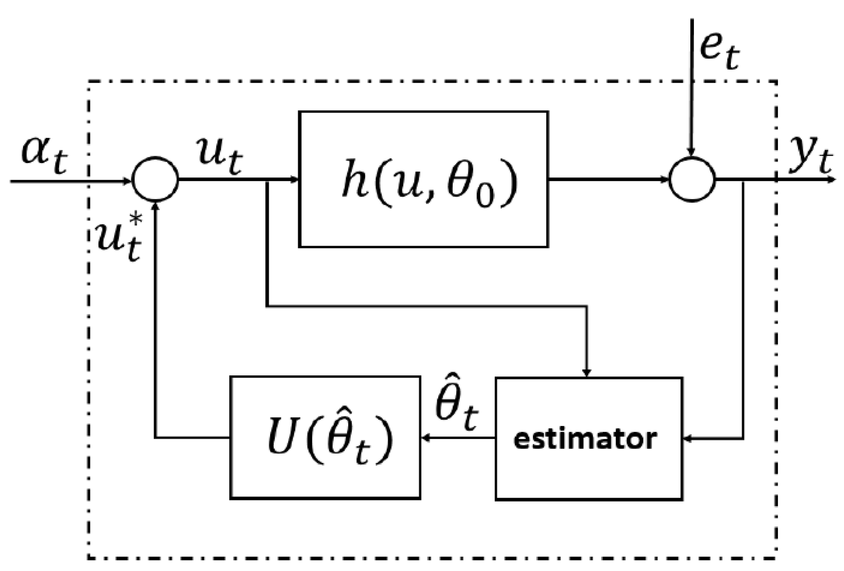
Challenge: The true parameter vector θ_0 is unknown

Certainty Equivalence Principle (CEP): Approximate u_0^* by replacing θ_0 with its estimate $\hat{\theta}_t$ learnt from the input $\{u_1,\dots,u_{t-1}\}$ and noisy measurement output $\{y_1,\dots,y_{t-1}\}$

$$u_t^* = \min_{u_t} \Phi(u,\hat{\theta}_t)$$

Problem: Due to the noise, u_t^* may not be informative enough to get an accurate estimate of θ_0

A dither-CEP framework



Input = Exploitation input + Exploration input

- Exploitation input $u_t^* = \min_{u_t} \Phi(u_t,\hat{\theta}_t)$: To take the best decision given the available information;
- Exploration input α_t: To get new information for an accurate parameter estimate.

However, α_t will reduce performance and thus there is a trade-off between exploitation and exploration when we design α_t.

Method

Problem: How to design an effective exploration strategy?
Method: Expected regret minimization

$$\min_{\substack{\alpha_t\\ t=1,\dots,T}} R_T = \min_{\substack{\alpha_t\\ t=1,\dots,T}} \sum_{t=1}^{T} \mathbb{E}[\underbrace{\Phi(u_t^*+\alpha_t,\theta_0)-\Phi(u_0^*,\theta_0)}_{\text{Instantaneous regret}}]$$

with $\alpha_t \sim N(0,x_t)$, where $x_t, t=1,\dots,T$ are to be designed.
Assumption: The estimator $\hat{\theta}_t$ for all t is unbiased and efficient.
Approximate regret dynamics (λ is a weight):

$$\tilde{R}_t = \tilde{R}_{t-1} + \mathbb{I}_t^{-1} + \lambda x_t\,, \quad t=1,\dots,T$$

Fisher Information dynamics:

$$\mathbb{I}_t = \mathbb{I}_{t-1} + \mathbb{E}[\frac{\partial h(u_t,\theta)}{\partial \theta}|^2_{u_t=u_t^*+\alpha_t}] = \mathbb{I}_{t-1} + f(x_t,\mathbb{I}_{t-1}^{-1})\,, \quad t=1,\dots,T$$

Nonlinear optimal control problem $\Longrightarrow$ optimal exploration

Assumption: The function f is non-negative, convex, and non-decreasing w.r.t. its arguments.
Definition: $x^* \in \mathbb{R}^T$ is a lazy excitation if $x_k^*=0$, for all k; x^* is an immediate excitation if $x_1^*>0$ and $x_k^*=0$, for $k\geq 2$.
Theorem: The optimal solution is a lazy or immediate excitation.

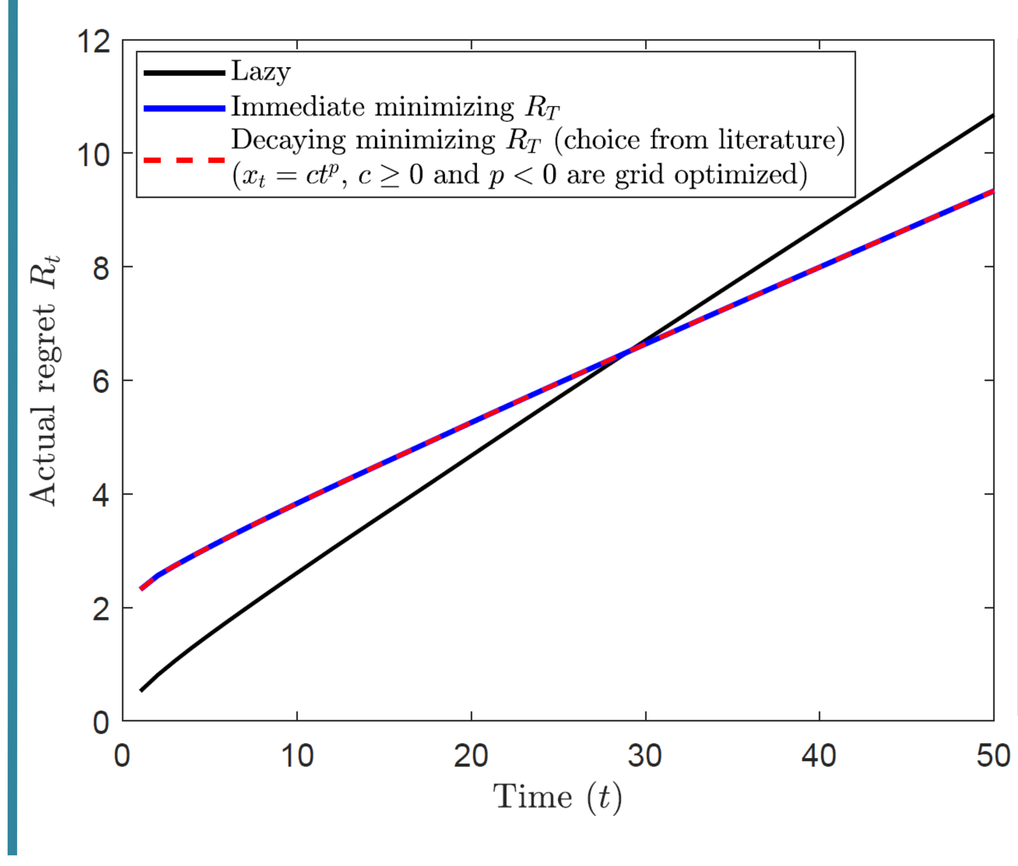
Example

The objective function and static input-output relationship are

$$\Phi(u,\theta_0) = u^2 + 2(\theta_0+1)u$$
$$y_t = \theta_0 u_t^2 + e_t, e_t \sim N(0,1)$$

where $\theta_0=-0.4$. The CEP exploration input is $u_t^* = -(\hat{\theta}_t+1)$. The oracle exploitation input, minimizing the cost, is $u_0^*=-0.6$. The exploration input $\alpha_t \sim N(0,x_t)$, where x_t, $t=1,\dots,T$ are to be designed by minimizing expected regret

$$\min_{\substack{x_t\\ t=1,\dots,T}} \tilde{R}_T = \min_{\substack{x_t\\ t=1,\dots,T}} \sum_{t=1}^{T} (\mathbb{I}_t^{-1} + x_t)$$
$$\mathbb{I}_t = \mathbb{I}_{t-1} + 3x_t^2 + [6\mathbb{I}_{t-1}^{-1} + 6(u_0^*)^2]x_t + 3\mathbb{I}_{t-1}^{-2} + 6(u_0^*)^2\mathbb{I}_{t-1}^{-1} + (u_0^*)^4$$



Analysis: When free information is enough, we can do nothing, as lazy excitation indicates. When exploration is necessary, it is best to explore it as early as possible since the reward, due to a better model, accumulates over the entire horizon T, rather than a part of it.

References

1. Colin, K., Hjalmarsson, H., & Bombois, X. (2022). Optimal exploration strategies for finite horizon regret minimization in some adaptive control problems. arXiv preprint arXiv:2211.07949.
2. Colin, K., Ferizbegovic, M., & Hjalmarsson, H. (2022). Regret Minimization for Linear Quadratic Adaptive Controllers Using Fisher Feedback Exploration. IEEE Control Systems Letters, 6, 2870-2875.
3. Wang, Y., Pasquini M., Colin, K., & Hjalmarsson, H. (2024). Regret Minimization in Scalar, Static, Non-linear Optimization Problems. Preprint at https://arxiv.org/abs/2403.15344.

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM