Reinforcement Learning in Control Theory

A New Approach to Mathematical Problem Solving

Kala Agbo Bidi [1] Jean-Michel Coron [1] Amaury Hayat [2] Nathan Lichtlé [2,3]

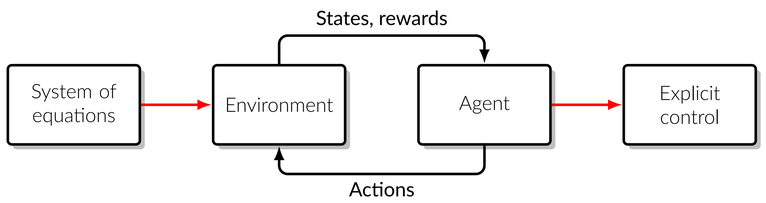
[1]Sorbonne University [2]École des Ponts ParisTech [3]UC Berkeley

NEURAL INFORMATION PROCESSING SYSTEMS

Feedback control stabilization using RL

Achieving stability through control feedback laws is one of the central questions in control theory. We present an approach mixing RL and mathematical analysis to solve this problem in complicated cases where mathematical theories are limited. We illustrate it on the SIT mosquito population model. Despite the mathematical complexities and absence of known solutions for this problem, our RL approach identifies an explicit candidate solution for a stabilizing control.

1. **Discretize the equations** in a numerical scheme and create a simulation from the dynamics
2. **Train an RL model** that learns to stabilize the system through many simulations
3. **Recover an explicit mathematical control** from the learned numerical control feedback
4. **Perform several tests** to ensure that the explicit control is efficient.



SIT model for pest control

Here we consider the case of **mosquitoes control**:

$$\begin{aligned}
\dot{E} &= \beta_E F\left(1-\frac{E}{K}\right) - (\nu_E+\delta_E)E && \leftarrow \text{mosquito density in aquatic phase}\\
\dot{M} &= (1-\nu)\nu_E E - \delta_M M && \leftarrow \text{wild adult male density}\\
\dot{F} &= \nu\nu_E E\frac{M}{M+M_s} - \delta_F F && \leftarrow \text{fecundated adult female density}\\
\dot{M}_s &= u - \delta_s M_s && \leftarrow \text{sterilized adult male density}
\end{aligned}$$

Control $u(t)$: sterilized adult males released at time t

Let $F_s(t) = F(t)M_s(t)/M(t)$ be the non-fecundated female density, then define:

$$\begin{aligned}
M_T(t) &= M(t) + M_s(t) && \leftarrow \text{total males}\\
F_T(t) &= F(t) + F_s(t) && \leftarrow \text{total females}
\end{aligned}$$

Mathematical open question: find $u(t) = f(M_T(t), F_T(t))$ where $f \in L^\infty(\mathbb{R}^2)$ such that the zero equilibrium $(0,0,0)$ is globally asymptotically stable and M_s is asymptotically small, i.e.

$$\lim_{t\to+\infty} \|E(t), M(t), F(t)\| = 0, \text{ and } \lim_{t\to+\infty} \|M_s(t)\| = \varepsilon, \tag{1}$$

where ε can be chosen arbitrarily small (with f depending on ε).

No solution is known for this problem when[1]

$$\varepsilon < U^* := \frac{K\beta_E\nu(1-\nu)\nu_E^2\delta_s}{4(\delta_E+\nu_E)\delta_F\delta_M}\left(1-\frac{\delta_F(\nu_E+\delta_E)}{\beta_E\nu\nu_E}\right)^2.$$

[1] Almeida, Luís, et al. "Optimal control strategies for the sterile mosquitoes technique", Journal of Differential Equations, 311, pp. 229–266 (2022).

Reinforcement learning framework

We learn a control $u(t) = \pi_\theta(s_t)$ that maximizes the expected return $\mathbb{E}\left[\sum_t \gamma^t r_t\right]$ with $\gamma \in (0,1)$.

Observations

We observe a combination of:

- Total males $M_T(t)$
- Total females $F_T(t)$

For better convergence, we provide states are at several scales then normalize them:

$$s_t = \left(\frac{\min(M_T(t),k)}{k}, \frac{\min(F_T(t),k)}{k}\right)_{k\in\mathcal{K}}$$

Reward function

We penalize the norm of the states:

$$r_t = c_1\|E(t), M(t), F(t)\|_2 + c_2(t)\|M_s(t)\|_2$$

Around the end of the horizon, we penalize $M_s(t)$ to encourage training convergence.

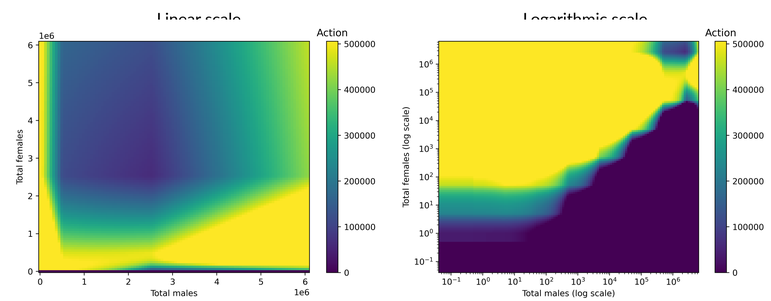
$$c_2(t) = \begin{cases} c_3 & \text{if } t < 0.9T,\\ c_3 + c_4 & \text{otherwise.}\end{cases}$$

Training algorithm

```
for iteration = 1, 2, ... do
    for rollout = 1, 2, ..., 12 do in parallel
        Generate a random initial condition (E, M, F, M_s)(0) ∈ [0, 50000]^4
        for D = 0, 1, ..., 1000 days do
            Get control action U(D) = π_θ(s_t) ∈ [0, 500000]
            Run 700 simulation steps (1 week) using u(t) = U(D)
            Compute reward r_t
        end for
    end for
    Optimize PPO loss[2] wrt. θ to maximize expected return using data collected during rollouts
    θ ← θ_optimized
end for
```

[2] Schulman, John, et al. "Proximal policy optimization algorithms", arXiv preprint arXiv:1707.06347 (2017).

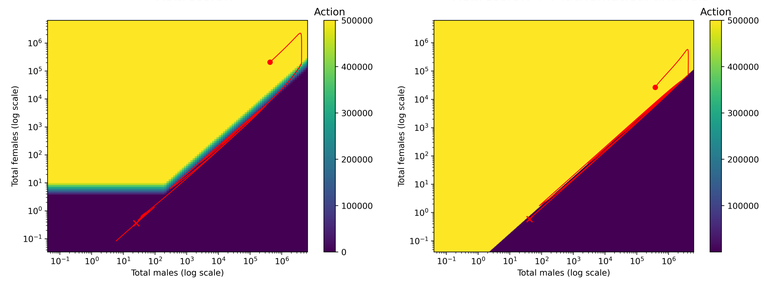
Learned control



The control first releases large amounts of sterilized male mosquitoes to make wild male and female populations decrease quickly, then carefully decreases the released amounts down to zero while making sure that the wild mosquito populations don't bounce back up.

From neural network to explicit control

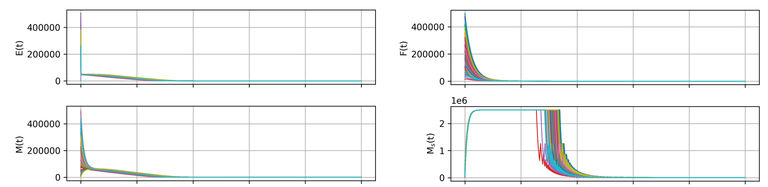
We fit the control in logarithmic scale since its behavior is much more apparent than in linear.



A first regression (left) more accurately fits the neural network's behavior but a simpler one (right) provides similar results, although it wouldn't have been obvious to come up with it:

$$\textbf{Final control:}\quad u_2(M_T, F_T) = \begin{cases} \varepsilon\delta_s & \text{if } \frac{\log(M_T)}{\log(F_T)} > \alpha_2,\\ u_{\max} & \text{otherwise,}\end{cases}$$

Using this control, $M_s(t)$ is large at first then quickly decreases as did other states:



Final control stats	200 days	400 days	600 days	800 days
average $\lvert E\rvert + \lvert M\rvert + \lvert F\rvert$	49×10^3	689	2.47	0.002
average $\lvert M_s\rvert$	2.5×10^6	129×10^3	2204	41.67

As we vary the ε in the control, we converge more or less quickly. The top plot is with a smaller ε, while the bottom plot is with a larger one. We can observe we release mosquitoes sporadically, less and less often, in order to prevent E, M and F from going back up while M_s decreases.