Reinforcement Learning for Control of Spatially Developing Distributed Systems

Amine Saibi[1,2], Onofrio Semeraro[2], Lionel Mathelin[2]

[1] Sorbonne Université, Inst. Jean Le Rond d'Alembert, 75005 Paris, France

[2] LISN, Université de Paris-Saclay, CNRS, 91400 Orsay, France

February 27, 2023

Supported by Agence National de la Recherche (ANR) under project ANR-21-CE46-0008 "Reinforcement Learning as Optimal control for Shear Flows REASON".

Summary

Let $V \subset \mathbb{R}^d$ be an open subset and $f : \mathbb{R}_+ \times V \times \mathbb{R}^m \longrightarrow \mathbb{R}^d$ be a continuous locally Lipschitz function.
For every ***control function*** $u \in \mathrm{L}^\infty(I, \mathbb{R}^m)$ we are interested in the Cauchy problem :

$$\begin{cases} \dot{x} = f(t, x, u), \\ x(0) = x_0, \end{cases} \tag{1}$$

with $x_0 \in V$.

Let $V \subset \mathbb{R}^d$ be an open subset and $f : \mathbb{R}_+ \times V \times \mathbb{R}^m \longrightarrow \mathbb{R}^d$ be a continuous locally Lipschitz function.
For every ***control function*** $u \in \mathrm{L}^\infty(I, \mathbb{R}^m)$ we are interested in the Cauchy problem :

$$\begin{cases} \dot{x} = f(t, x, u), \\ x(0) = x_0, \end{cases}$$

with $x_0 \in V$.
A solution associated to the control depending upon initial conditions is a function

$$\begin{aligned} x_u : \Omega &\longrightarrow \mathbb{R}^d \\ (x_0, t) &\longmapsto x_u(t; x_0). \end{aligned}$$

With $\Omega = \bigcup_{x_0} \Omega_{x_0}$ where $\Omega_{x_0} = \{x_0\} \times I_{x_0}$ and I_{x_0} is a maximal interval on which the solution is defined.

Let $V \subset \mathbb{R}^d$ be an open subset and $f : \mathbb{R}_+ \times V \times \mathbb{R}^m \longrightarrow \mathbb{R}^d$ be a continuous locally Lipschitz function.
For every ***control function*** $u \in \mathrm{L}^\infty(I, \mathbb{R}^m)$ we are interested in the Cauchy problem :

$$\begin{cases} \dot{x} = f(t, x, u), \\ x(0) = x_0, \end{cases}$$

with $x_0 \in V$.
A solution associated to the control depending upon initial conditions is a function

$$\begin{aligned} x_u : \Omega &\longrightarrow \mathbb{R}^d \\ (x_0, t) &\longmapsto x_u(t; x_0). \end{aligned}$$

With $\Omega = \bigcup_{x_0} \Omega_{x_0}$ where $\Omega_{x_0} = \{x_0\} \times I_{x_0}$ and I_{x_0} is a maximal interval on which the solution is defined.
Suppose also having two continuously differentiable functions $g : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}$ and $r : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^m \longrightarrow \mathbb{R}$ that define a ***cost function***:

$$\mathcal{J}(x_0, u, T) \coloneqq g(T, x_u(T)) + \int_0^T r(s, x_u(s), u(s))\mathrm{d}s,$$

where $]0, T[\subset I_{x_0}$ and x_u is the solution of (1) associated to a control $u \in \mathrm{L}^\infty$.

The optimal control problem consists in finding a control $u \in \mathrm{L}^\infty$ ***that minimizes*** $\mathcal{J}$.

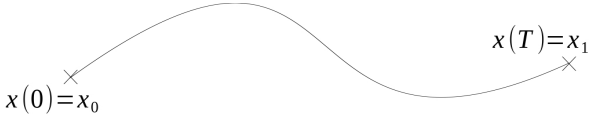
Considering an observed control system given initial condition $x(0) = x_0 \in \mathbb{R}^d$

$$\begin{cases} \dot{x} = f(t, x, u), \\ y = h(x), \end{cases} \tag{1}$$

with $h : \mathbb{R}^d \to \mathbb{R}^p$, an observation function.

Definition (Controllability)

We say that the dynamic (1) is ***controllable*** (in time T) when for every $x_0, x_1 \in \mathbb{R}^d$ there exists a control function $u \in \mathrm{L}^\infty$ such that the associated solution x_u verifies $x_u(0) = x_0$ and $x_u(T) = x_1$.



Definition (Observability)

We say that the dynamic (1) is ***observable*** when for every $x_0, x_1 \in \mathbb{R}^d$, if $x_0 \neq x_1$ then there exists a control function $u \in \mathrm{L}^\infty$ such that $y_u^{(0)} \neq y_u^{(1)}$.
With $y_u^{(0)}$ and $y_u^{(1)}$ the outputs associated with the solutions with initial conditions x_0 and x_1 respectively.

Summary

An important particular case is when the system is linear,

$$\begin{cases} \dot{x} = Ax + Bu + \nu, \\ y = Cx + \omega, \end{cases}$$

with Gaussian white noises ν and ω having covariance matrices $V \in \mathcal{M}_d(\mathbb{R})$ and $W \in \mathcal{M}_p(\mathbb{R})$ respectively, and given data $A \in \mathcal{M}_d(\mathbb{R})$, $B \in \mathcal{M}_{d,m}(\mathbb{R})$ and $C \in \mathcal{M}_{d,p}(\mathbb{R})$.

Together with a quadratic cost

$$\mathcal{J}(u) = \int_0^T \langle Qx(t)\,,\, x(t)\rangle + \langle Ru(t)\,,\, u(t)\rangle \,\mathrm{d}t,$$

where Q is positive semi-definite matrix and R a positive-definite one.

When the LQ problem is controllable and observable, a ***dynamic state feedback*** stabilizing control is given by

$$u(t) = K\hat{x}(t),$$

where $\hat{x}$ is an estimated state following the dynamic

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Bu + LC\,(x - \hat{x}), \\ \hat{x}(0) = \mathbb{E}[x_0]. \end{cases}$$

When the LQ problem is controllable and observable, a ***dynamic state feedback*** stabilizing control is given by

$$u(t) = K\hat{x}(t),$$

where $\hat{x}$ is an estimated state following the dynamic

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Bu + LC\,(x - \hat{x}), \\ \hat{x}(0) = \mathbb{E}[x_0]. \end{cases}$$

with $K = -R^{-1}B^T P$ and $L = GC^T W^{-1}$ where P and G are solutions of the Riccati equations

$$\begin{cases} \dot{P} + A^T P + PA - PBR^{-1}B^T P + Q = 0, \\ P(T) = 0. \end{cases} \qquad \begin{cases} \dot{G} - AG - GA^T + GC^T W^{-1} CG - V = 0, \\ G(0) = \mathbb{E}\left[x_0^T x_0\right]. \end{cases}$$

[1] Trélat, E. (2005). Contrôle optimal: théorie & applications. Paris: Vuibert.

Summary

Let $\left(\mathcal{X}, \mathcal{F}^{\mathcal{X}}\right)$ and $\left(\mathcal{A}, \mathcal{F}^{\mathcal{A}}\right)$ be two measurable spaces that represent respectively the state space and the action space.
Consider given:

$$P : \mathcal{X} \times \mathcal{A} \longrightarrow [0,1]^{\mathcal{F}^{\mathcal{X}}} \qquad \pi : \mathcal{X} \longrightarrow [0,1]^{\mathcal{F}^{\mathcal{A}}}$$
$$(x,a) \longmapsto P(x,a,\cdot) \qquad x \longmapsto \pi(x,\cdot)$$

such that:

- $\forall (x,a) \in \mathcal{X} \times \mathcal{A}$: $P(x,a,\cdot)$ is a probability measure called transition law.
- $\forall x \in \mathcal{X}$: $\pi(x,\cdot)$ is a probability measure called the policy.

This, provided an initial state-action pair (x_0, a_0), defines a pair of stochastic ***state-action*** processes $(X_t, A_t)_{t \in \mathbb{N}}$ with values in $\mathcal{X} \times \mathcal{A}$.

Let $\left(\mathcal{X}, \mathcal{F}^{\mathcal{X}}\right)$ and $\left(\mathcal{A}, \mathcal{F}^{\mathcal{A}}\right)$ be two measurable spaces that represent respectively the state space and the action space.
Consider given:

$$P : \mathcal{X} \times \mathcal{A} \longrightarrow [0,1]^{\mathcal{F}^{\mathcal{X}}} \qquad \pi : \mathcal{X} \longrightarrow [0,1]^{\mathcal{F}^{\mathcal{A}}}$$
$$(x,a) \longmapsto P(x,a,\cdot) \qquad x \longmapsto \pi(x,\cdot)$$

such that:

- $\forall (x,a) \in \mathcal{X} \times \mathcal{A}$: $P(x,a,\cdot)$ is a probability measure called transition law.
- $\forall x \in \mathcal{X}$: $\pi(x,\cdot)$ is a probability measure called the policy.

This, provided an initial state-action pair (x_0, a_0), defines a pair of stochastic ***state-action*** processes $(X_t, A_t)_{t \in \mathbb{N}}$ with values in $\mathcal{X} \times \mathcal{A}$.

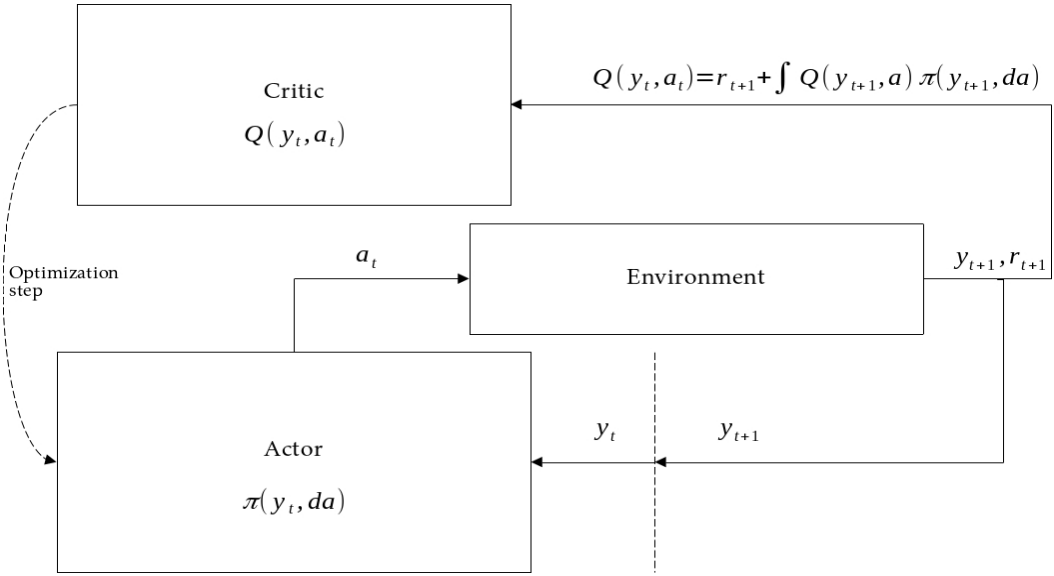
Define also a ***gain function*** $g : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \longrightarrow \mathbb{R}$ and the ***reward process*** :

$$R_{t+1} = g(X_t, A_t, X_{t+1}).$$

We call value function the expected cumulated rewards :

$$Q_\pi(x_0, a_0) = \mathbb{E}_{\mathbb{P}_{x_0,a_0,P,\pi}}\left[\sum_{t \geq 1} R_t\right].$$

[1] Puterman, M.L. (2014). Markov Decision Processes : Discrete Stochastic Dynamics. John Wiley & Sons.



[1] Lillicrap T.P. et al. (2015), Continuous control with deep reinforcement learning, arXiv.

Alternatively, one can approximate an unknown dynamic from input/output data $\{u_k, y_k\}_{k\in\{0,..,n\}}$ with linear control system, i.e $\hat{A}$, $\hat{B}$ and $\hat{C}$ such that for $k \in \{0,..,n\}$

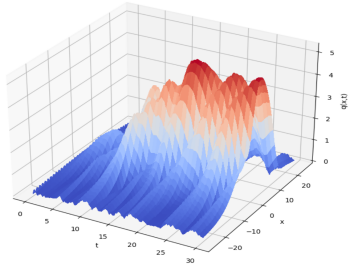
$$\begin{cases} x_{k+1} = \hat{A}x_k + \hat{B}u_k + \nu_k, \\ y_k = \hat{C}x_k + \omega_k, \end{cases}$$

fits the data.

We achieve this using ***subspace methods*** (N4SID), then build an LQG design for the resulting system.

[1] Van Overschee, P., & De Moor, B. (2012). Subspace identification for linear systems. Springer Science & Business Media.

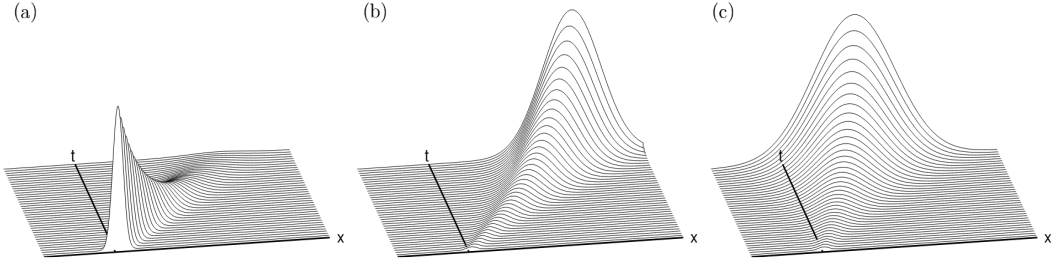
Summary



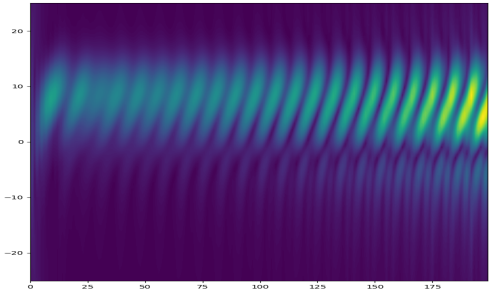
Complex Ginzburg-Landau equation

$$q : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{C}$$
$$(t, x) \longmapsto q(t, x).$$

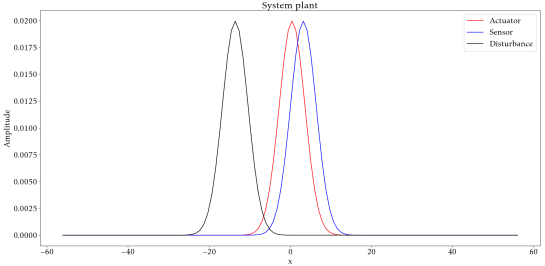
$$\partial_t q = \gamma \partial_x^2 q + \nu \partial_x q + \mu q + \alpha |q|^2 q.$$



[1] Bagheri, S., Henningson, D. S., Hoepffner, J., & Schmid, P. J. (2009). Input-output analysis and control design applied to a linear model of spatially developing flows. Applied Mechanics Reviews, 62(2).



$$\partial_t q = \gamma \partial_x^2 q + \nu \partial_x q + \mu q + \alpha |q|^2 q + F.$$



Control forcing

$$\begin{cases} F(t,x) = \langle B(x)\,,\, u(t) \rangle, \\ B_j(x) = e^{-\frac{(x-a_j)^2}{\sigma^2}}. \end{cases}$$

Sensor measures

$$\begin{cases} y(t) = \langle C(\cdot)\,,\, q(t,\cdot) \rangle_{\mathrm{L}^2}, \\ C_i(x) = e^{-\frac{(x-s_i)^2}{\sigma^2}}. \end{cases}$$

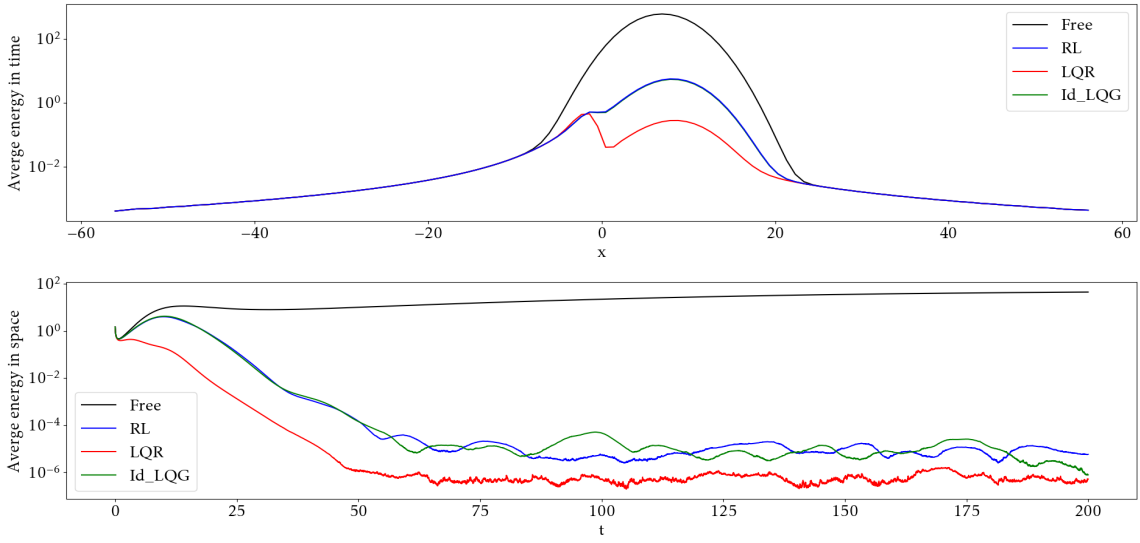


Figure: Energy curves for sensor position $x_s = 2.5$, actuator position $x_a = 0$ and disturbance at $x_d = -14.0$.

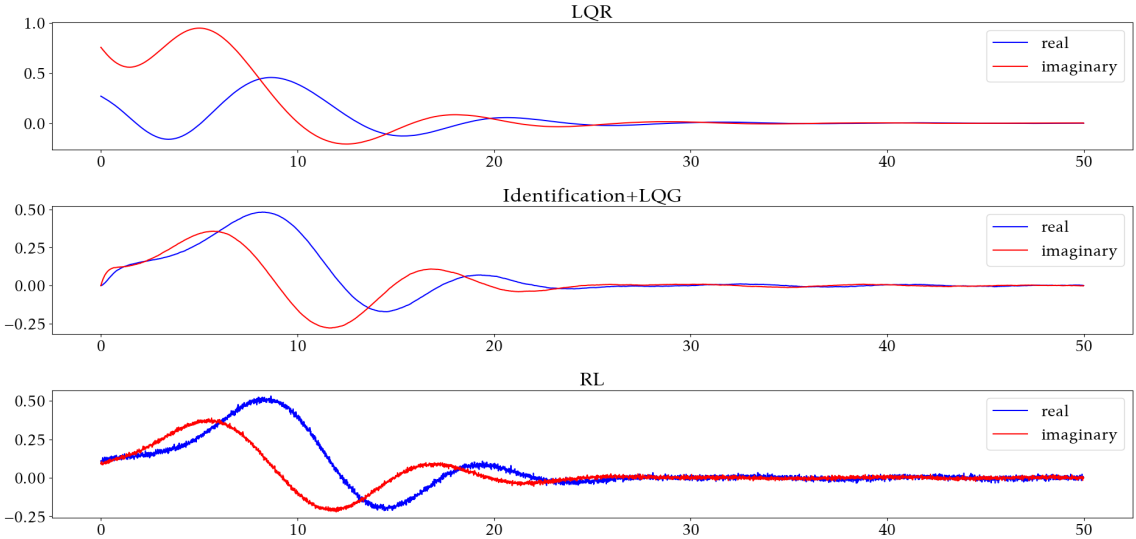


Figure: Control signals for sensor position $x_s = 2.5$, actuator position $x_a = 0$ and disturbance at $x_d = -14.0$.

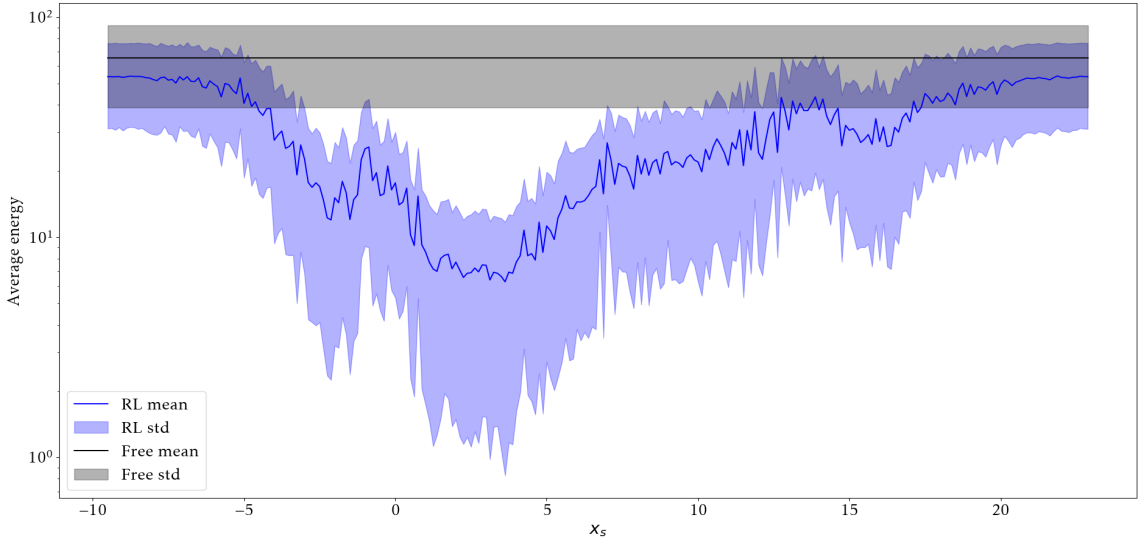
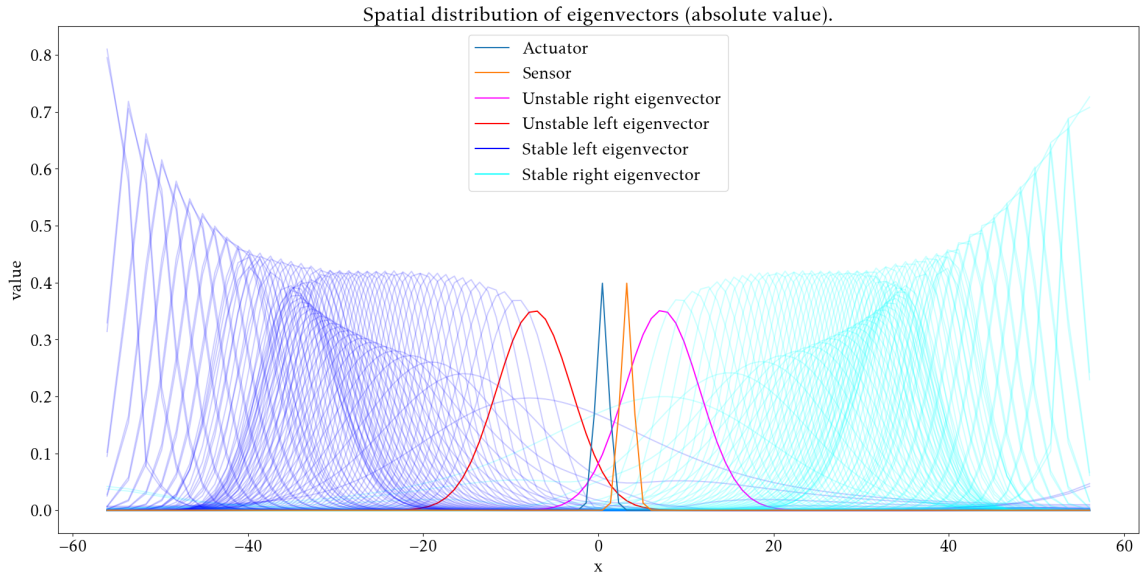


Figure: At each x_s, five models are trained with different noise seeds, and tested with five other noise seeds, the performance is averaged on those seeds.


Spatial distribution of eigenvectors (absolute value).
Actuator
Sensor
Unstable right eigenvector
Unstable left eigenvector
Stable left eigenvector
Stable right eigenvector
value
x
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
−60
−40
−20
0
20
40
60

Concluding remarks and observations

- In case of observable/controllable dynamics, one can with few data learn effective control strategies in a model-free fashion.
- For our test cases, the (sub-optimal) strategies obtained show robustness w.r.t changes in the dynamic's parameters and stochastic properties of the noises.
- Control laws obtained are "reasonable" in view of the optimal ones.
- Limited theoretical marginal garantees in reinforcement learning.