FLEX: an Adaptive Exploration Algorithm for Nonlinear Systems

Matthieu Blanke, Marc Lelarge

matthieu.blanke@inria.fr marc.lelarge@inria.fr

Inria & ENS, Paris

ICML 2023

Paper and project page :

SCAN ME

The exploration problem

Control and reinforcement learning

Pyrène robot, credit : CNRS

Airbus A320neo, credit : Airbus

An accurate control model requires collecting experimental data, which is costly.

Objective (Exploration)

Navigate the system towards informative states for efficient learning.

Controlled dynamics

The mathematical model

Controlled, unknown dynamical system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f_\star(x, u)$$

observe noisy $x \in \mathbb{R}^d$, learn unknown $f_\star$ with $f(x, \theta)$, choose $u \in \mathbb{R}^m$

Algorithm Active exploration

input model f, policy π, time horizon T, time step $\mathrm{d}t$, estimator $\hat{\theta}$
output model estimate θ_T
for $0 \leq t \leq T-1$ **do**
  choose $u_t = \pi(x_{0:t}, u_{0:t-1}; \theta_t)$
  observe $x_{t+1} = x_t + \mathrm{d}t\, f_\star(x_t, u_t)$ + noise
  update $\theta_{t+1} = \hat{\theta}(x_{0:t+1}, u_{0:t})$
end for

Objective (Exploration)

$$\min_{\pi} \quad \mathbb{E}\left[\|f_\star - f(., \theta_T)\| | \pi\right]$$

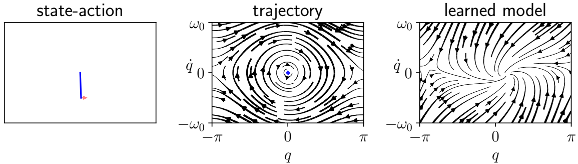
Find an exploration policy π that yields **informative trajectories** (x_t) for the model f.

Requirement The policy should run online, in real time.

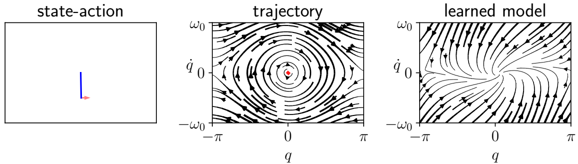
An illustrative example

The pendulum

Random



Resonant



FLEX

An adaptive exploration algorithm

Related work (Modeling uncertainty for nonlinear dynamics)

Gaussian processes [Buisson-Fenet *et al.*, 2020], ensemble of neural nets [Shyam *et al.*, 2019], Random Fourier Features [Schultheis *et al.* 2019]. Computationally heavy, offline planning.

Based on information theory, our policy maximizes the Fisher information about the model.

Result (Online D-optimal exploration)

One-step-ahead information maximization can be solved with a quadratic program :

$$\max_{u \in \mathbb{R}^m} \quad u^\top Q_t u - 2 v_t^\top u$$
$$\text{subject to} \quad u^\top u = \gamma^2.$$

▷ Can be solved fast and online.

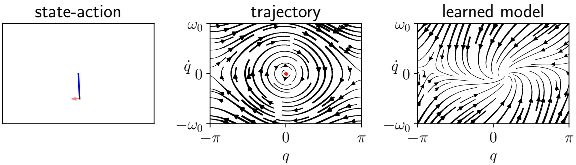
Algorithm FLexible EXploration (FLEX)

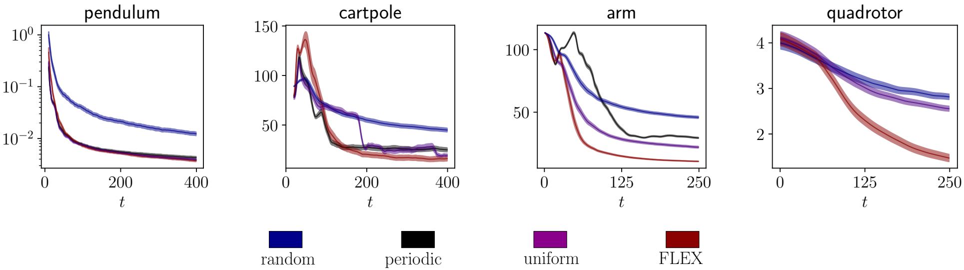
input model f, time horizon T
output parameter estimate θ_T
for $0 \leq t \leq T-1$ **do**
  choose $u_t \in \operatorname{argmax}_{\|u\|^2 \leq \gamma^2} u^\top Q_t u - 2 v_t^\top u$
  observe $x_{t+1} = x_t + \mathrm{d}t\, f_\star(x_t, u_t) + \text{noise}$
  online learning $\theta_{t+1} = \text{update}(\theta_t, x_{0:t+1})$
end for

Experimental benchmark

Model error over time

FLEX



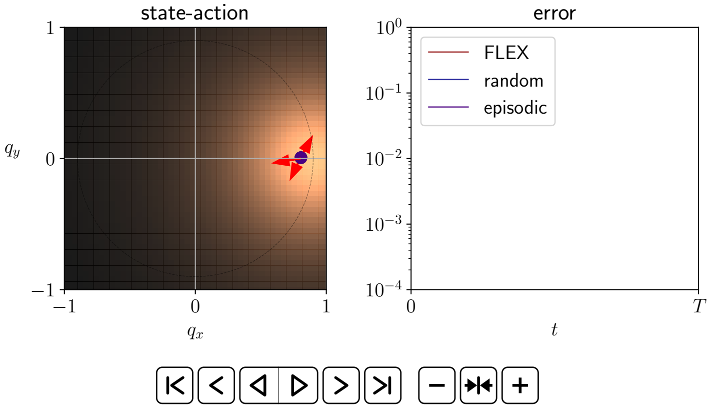


Evaluating with exploitation

Sample complexity for the swing-up task

Method	random	MAX	SAC	RHC	FLEX
pendulum					
samples	✗	2000	✗	500	50
compute	1	100	2	8	4
cartpole					
samples	✗	✗	✗	600	300
compute	1	20	1.5	2	1.6

Tracking time-varying dynamics
Adaptive



Check out the paper

Paper and project page :

SCAN ME