Exploration of physical systems

INRIA Junior Seminar

Matthieu BLANKE

INRIA Paris, team ARGO

October 18th, 2022

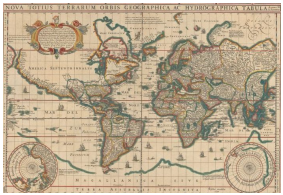
Exploration and physical systems

Exploration of physical systems

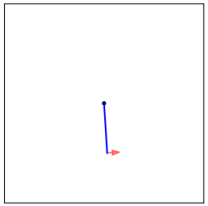
Exploration : an agent learns an unknown environment by moving and searching information.

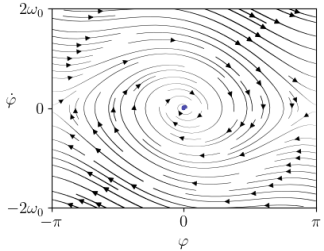
Brierly, O. W., iscovery of the Straits of Magellan in 1520.

Frederik De Wit's 1654 Dutch Sea Atlas. Image courtesy of the Harvard Map Collection.



The damped simple pendulum and its phase portrait.

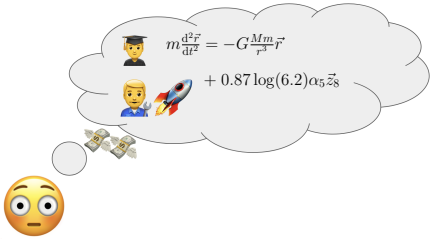




Motivation : control and reinforcement learning

Exploration of physical systems

source : NASA



Optimal control theory has led to great successes, but needs a faithful model of the system.
To this end, experimental measurements can complete well-established theory.
Running experiments is costly : need for efficiency.

Motivation

Control the system so that the model is learned efficiently from the experiment.

Organization of the presentation

Exploration of physical systems

1. Problem formulation
2. Linear system identification
3. Nonlinear exploration

Controlled dynamics

Problem formulation

Controlled dynamical system $\boxed{\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,u)}, \quad x \in \mathbb{R}^d, \quad u \in \mathbb{R}^m, \quad$ unknown f,

from which we collect observations $\boxed{x_{t+1} = x_t + \mathrm{d}t f(x_t, u_t) + w_t, \quad 0 \leq t \leq T-1}$,

with noise $w_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 I_d)$. The input is chosen with the magnitude constraint $\boxed{\|u_t\|_2 \leq \gamma}$.

Learning rule

Given a trajectory, the dynamics f are learned with a parametric model f_θ by regression :

$$\theta_t \in \underset{\theta \in \mathbb{R}^q}{\operatorname{argmin}} \frac{1}{2} \sum_{s=0}^{t-1} \|(x_{s+1} - x_s)/\mathrm{d}t - f_\theta(z_s)\|_2^2, \quad \text{with} \quad z_s = \begin{pmatrix} x_s \\ u_s \end{pmatrix} \in \mathbb{R}^{d+m}.$$

Objective (informal)

Our aim is to choose (u_t) so that f is learned as fast as possible from the trajectory (z_t).

Example : damped pendulum

Problem formulation

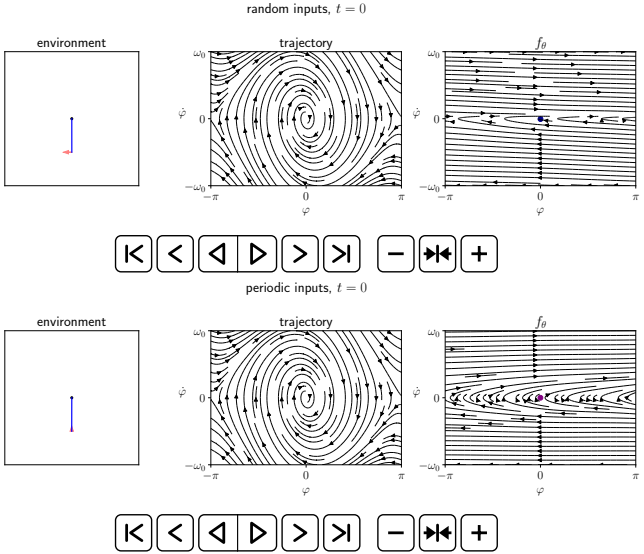
Damped pendulum :

$$d = 2,\ m = 1,\ x = \begin{pmatrix} \varphi \\ \dot{\varphi} \end{pmatrix}.$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \varphi \\ \dot{\varphi} \end{pmatrix} = \begin{pmatrix} \dot{\varphi} \\ -\alpha\dot{\varphi} - \omega_0^2 \sin\varphi + \beta u \end{pmatrix} := f(\varphi, \dot{\varphi}, u)$$

Baselines

- Random inputs : $u_t \sim \mathcal{N}(0, \frac{\gamma^2}{m} I_d)$
- Periodic inputs : $u_t = \gamma \sin \omega_0 t$



Objective

Problem formulation

Objective

Given a parametric model f_θ and a fixed learning rule $\hat{\theta} : (x_{0:t}, u_{0:t-1}) \mapsto \theta_t \in \mathbb{R}^q$, find a **policy** $\pi : (x_{0:t}, u_{0:t-1}) \mapsto u_t$ that yields **informative trajectories** (z_t) for the model f_θ.

Requirements

- ▷ Sample-efficiency : at time T, $f_\theta \simeq f$.
- ▷ Adaptability : update π with θ_t at each t .
- ▷ Speed : computing u_t must be fast.

Algorithm Active exploration

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

Question How to measure informativeness ?

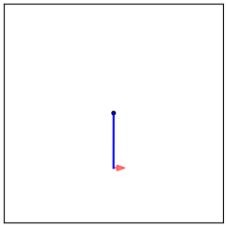
Organization of the presentation

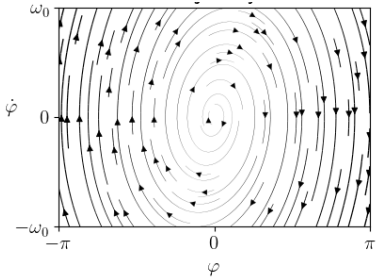
Exploration of physical systems

1. Problem formulation
2. Linear system identification
3. Nonlinear exploration

Motivation

- Computations are easier.
- Linear system are sufficient to describe systems near a stable equilibrium (LQR control).
- When linearity does not apply globally, it still applies locally.





Linearized damped pendulum :

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\varphi\\ \dot{\varphi}\end{pmatrix} \simeq \begin{pmatrix}\dot{\varphi}\\ -\alpha\dot{\varphi} - \omega_0^2\varphi + \beta u\end{pmatrix}.$$

Setting

Linear system identification

The flow f is a linear function of x and u :

$$\boxed{\begin{array}{c} x_{t+1} = A_* x_t + B_* u_t + w_t, \quad 0 \leq t \leq T-1, \\ x_0 = 0. \end{array}}$$

Linear relation $x_{t+1} = \theta_\star \times z_t + w_t$ with $\theta_* = (A_* \, B_*) \in \mathbb{R}^{d \times (d+m)}$ and $z_t = \begin{pmatrix} x_t \\ u_t \end{pmatrix}$.

Learning rule (Linear dynamics)

The squared error $\ell(\theta) = \dfrac{1}{2\sigma^2} \displaystyle\sum_{t=0}^{T-1} \|x_{t+1} - Ax_t - Bu_t\|_2^2$ is minimized by

the **least squares estimator** $\boxed{\hat{\theta}_T = {M_T}^{-1} \displaystyle\sum_{t=0}^{T-1} z_t x_{t+1}^\top}$ with $\boxed{M_t = \displaystyle\sum_{s=0}^{t-1} z_s z_s^\top}$ the Gram matrix.

Linear optimal design

Linear system identification

Theory of **optimal experimental design** : for least squares, the signal is maximally informative if the eigenvalues of M_T are "maximized" (see [Pukelsheim, 2006] for details).

Typically,

$$\begin{aligned} \max \quad & \det\left(\sum_{t=0}^{T-1} z_t z_t^\top\right) \\ \text{subject to} \quad & x_{t+1} = A_\star x_t + B_\star u_t + w_t \\ & \|u_t\|_2 \leq \gamma \quad \forall t. \end{aligned}$$

D-optimal inputs

Related work (Optimal design for linear dynamical systems)

- Theory of optimal design for control in the 70s : [Mehra 1976, Goodwin1977].
- Interest from the machine learning community lately, providing theoretical asymptotic bounds [Simchowitz *et al.* 2018] and algorithms [Wagenmaker & Jamieson 2021] (TOPLE algorithm).

Greedy identification

Linear system identification

Result (B. & Lelarge, CDC 2022)

One-step-ahead D-optimal planning

$$\max_{\|u_t\|^2 \leq \gamma^2} \log \det \left(M_t + \mathbb{E}[z_{t+1} z_{t+1}^\top | \theta_t] \right)$$

takes the form

$$\begin{aligned} \max_{u \in \mathbb{R}^m} \quad & u^\top Q_t u - 2 v_t^\top u \\ \text{such that} \quad & u^\top u = \gamma^2 \end{aligned}$$

Efficient numerical solution, at the cost of a $d \times d$ matrix eigendecomposition for the computation of Q_t and v_t and a scalar root search.

Algorithm Greedy system identification

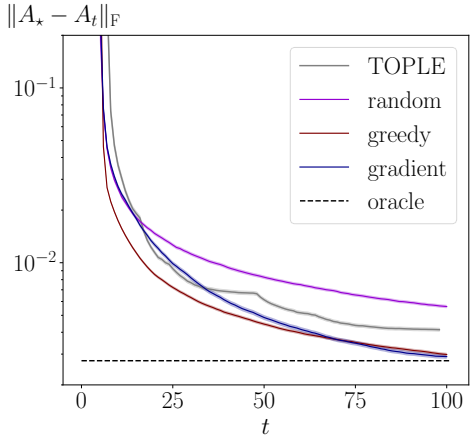
output final estimate θ_T
for $0 \leq t \leq T-1$ **do**
  $u_t \in \operatorname{argmax}_{\|u\|^2 \leq \gamma^2} u^\top Q_t u - 2 v_t^\top u$
  play u_t, observe x_{t+1}
  $M_t = M_t + z_{t+1} z_{t+1}^\top$
  $\theta_{t+1}^\top = M_{t+1}^{-1} \left(M_t \theta_t^\top + z_t x_{t+1}^\top \right)$
end for

The algorithm is **adaptive** and **fast**.
Is it **sample efficient** ?

Identification performance

Linear system identification

Set-up : $d = 4$, random $\theta_\star = A_\star$, $B = I_d$, $\gamma = 1$, $\sigma = 0.1$, $T = 100$



Realistic system : lateral motion of a Lockheed Jet star. Numerical values from a report. Here, $T = 125$ for the test flight.

Policy	error	time
Random	1.1×10^{-1}	1
TOPLE	8.6×10^{-2}	~ 50
gradient	8.3×10^{-2}	~ 25
Greedy	8.2×10^{-2}	~ 2
Oracle	8.0×10^{-2}	~ 100

Organization of the presentation

Exploration of physical systems

1. Problem formulation
2. Linear system identification
3. Nonlinear exploration

From linear to nonlinear systems

Nonlinear exploration

Which parametric model f_θ for a nonlinear map ?

Related work

Natural in Bayesian models like Gaussian processes [Buisson-Fenet *et al.*, 2020], ensemble of neural nets [Shyam *et al.*, 2019], Random Fourier Features [Schultheis *et al.* 2019].

We'd like to generalize to nonlinear dynamics, keeping the algorithm **online** and **fast**.
Neural network-based models for nonlinear maps are appealing for their expressive power.

Learning rule (Nonlinear dynamics)

We extend online least squares with **online gradient descent** :

$$\ell_t = \|f_\theta(x_t, u_t) - (x_{t+1} - x_t)/\mathrm{d}t\|_2^2, \quad \theta_{t+1} = \theta_t - \eta \nabla \ell_t(\theta_t).$$

Linearized optimal design

Nonlinear exploration

Optimal design is suited for **linear parametrizations**. The optimization problem was simple because the **dynamics were linear**.

Approximations

▷ Linearize the model **with respect to the parameters** [MacKay 1992] :

$$f_\theta(z,\theta) \simeq f_\theta(z,\theta_\star) + J_\theta(z,\theta_\star) \times (\theta - \theta_\star) \quad \text{with} \quad J_\theta(z,\theta) := \frac{\partial f_\theta}{\partial \theta}(z,\theta) \in \mathbb{R}^{d\times q}.$$

▷ Linearize the dynamics **with respect to the state** :

$$J_\theta(x,\theta) \simeq J_\theta(x_t,\theta) + \frac{\partial J_\theta}{\partial x}\mathrm{d}x \quad \text{and} \quad \mathrm{d}x \simeq \frac{\partial f_\theta}{\partial u}\,\mathrm{d}t\,u.$$

Idea : compute $B_t = \dfrac{\partial f_\theta}{\partial u}$, $D_t = \dfrac{\partial^2 f_\theta^{(k)}}{\partial x\,\partial\theta} \in \mathbb{R}^{d\times q}$. $J_t = J_\theta(z_t,\theta_t)$, $M_t = \displaystyle\sum_{s=0}^{t-1} {J_s}^\top J_s \in \mathbb{R}^{q\times q}$

evaluated at z_t and generalize the computations of the linear case.

Linearized D-optimal exploration

Nonlinear exploration

Result

Linearized D-optimal planning

$$\max_{\|u_t\|^2 \leq \gamma^2} \log \det \Big(M_t + \mathbb{E}[J_t^\top J_t | \theta_t] \Big)$$

can be approximated at first order in $\gamma \mathrm{d}t$ by

$$\begin{aligned} \max_{u \in \mathbb{R}^m} \quad & u^\top Q_t u - 2 v_t^\top u \\ \text{such that} \quad & u^\top u = \gamma^2 \end{aligned}$$

where Q_t and v_t are computed with B_t and D_t.

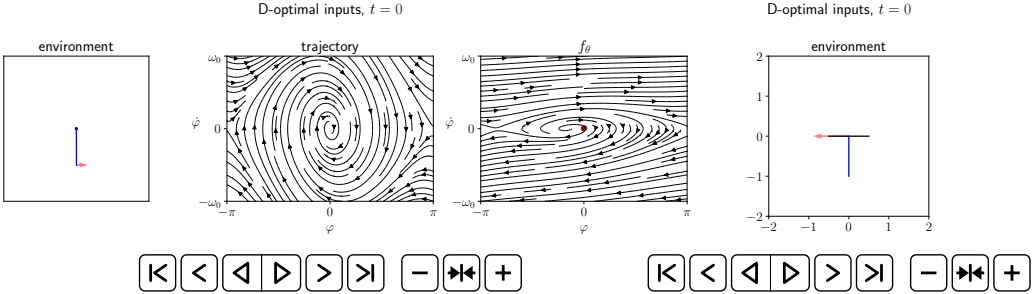
The cost dominated by the computation of the second derivatives in D.

Algorithm Nonlinear exploration

input neural model f_θ, policy π, time horizon T, time-step $\mathrm{d}t$, learning rate η
output dynamics model f_θ
for $0 \leq t \leq T-1$ **do**
  choose $u_t \in \underset{\|u\|^2 \leq \gamma^2}{\operatorname{argmax}} \, u^\top Q_t u - 2 v_t^\top u$
  observe $x_{t+1} = x_t + \mathrm{d}t \, f(x_t, u_t)$
  compute the loss
    $\ell_t = \| f_\theta(x_t, u_t) - (x_{t+1} - x_t)/\mathrm{d}t \|_2^2$
  update $\theta \leftarrow \theta - \eta \nabla \ell_t(\theta)$
end for

D-optimal trajectories

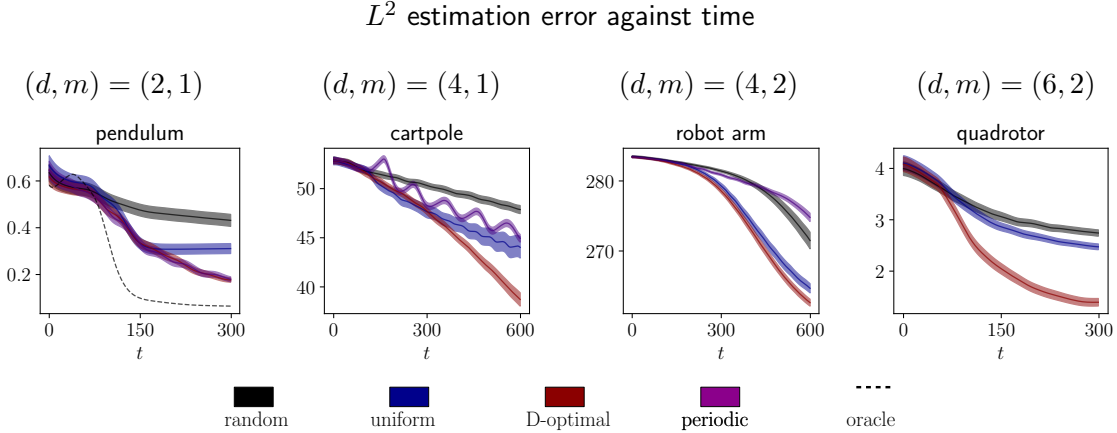
Nonlinear exploration



Experimental benchmark

Nonlinear exploration

L^2 estimation error against time



Conclusion

• We devised an online exploration algorithm for linear and nonlinear dynamics, based on information-theoretic computations.

• No theoretical guarantees but good performance in a practical framework.

• Prospectives : measure performances in terms of a downstream exploitation task ? scaling to larger models ?

Questions ?