Reading group 1
Control systems and RL

Antoine Bambade

Willow and Sierra teams

Contents

01

Introduction

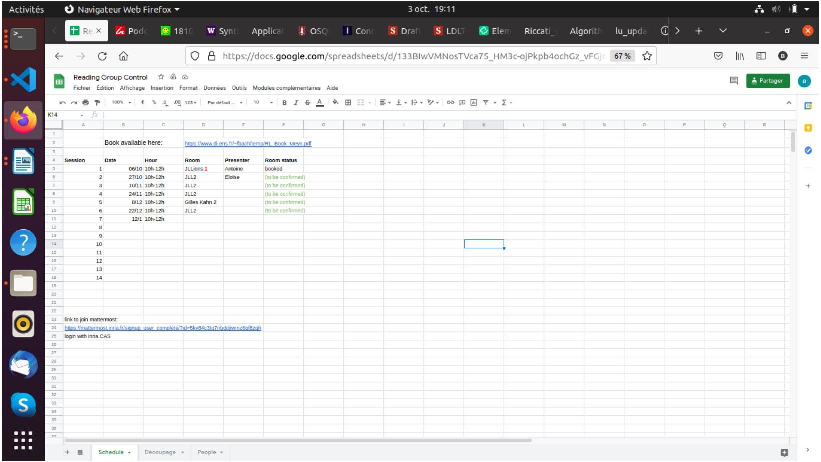
Objective

- Discover or deepen RG topic

Support

- "Control systems and Reinforcement Learning", Sean Meyn
- Discussion
- papers
- talks
- ...

Organization



Reading Group Control

Book available here: https://www.di.ens.fr/~fbach/temp/RL_Book_Meyn.pdf

Session	Date	Hour	Room	Presenter	Room status
1	06/10	10h-12h	JLLions 1	Antoine	booked
2	27/10	10h-12h	JLL2	Eloïse	(to be confirmed)
3	10/11	10h-12h	JLL2		(to be confirmed)
4	24/11	10h-12h	JLL2		(to be confirmed)
5	8/12	10h-12h	Gilles Kahn 2		(to be confirmed)
6	22/12	10h-12h	JLL2		(to be confirmed)
7	12/1	10h-12h			
8					
9					
10					
11					
12					
13					
14					

link to join mattermost:
https://mattermost.inria.fr/signup_user_complete/?id=5ky84c3tq7nbddjjwmz6qf8zqh
login with inria CAS

02

Control "Crash Course"

Plan

- 1) Notations
- 2) What to do ?
  - modeling
  - assessing performance
- 3) RL "main" algorithms

Notations

- *ff* : feedforward control
- u : inputs
- y : observations
- ϕ : policy

Parameters

- *ff* : feedforward control
- u : inputs
- y : observations
- ϕ : policy

Often : $u(k) = u_{ff}(k) + u_{fd}(k)$

Parameters

- ff : feedforward control
- u : inputs
- y : observations
- ϕ : policy

Often : $u(k) = u_{ff}(k) + u_{fd}(k)$

What to do about it?

- 1) Ceate model

$$y(k) = G_k(u(0), u(1), ..., u(k)), k \geq 0$$

- 2) Design policy ϕ
- 3) Assess performance

Objective

Design a **simple** and **faithful** model.

$$y(k) = G_k(u(0), u(1), ..., u(k)), k \geq 0$$

Examples of models

- Linear and Time invariant (LTI)

$$y(k) = \sum_{i=0}^{k} b_i u(k-i), k \geq 0$$

- Auto-Regressive Moving-Average (ARMA)

$$y(k) = -\sum_{i=1}^{N} a_i y(k-i) + \sum_{i=0}^{M} b_i u(k-i)$$

Definition

- State space $\mathbf{X}$
- Action space $\mathbf{U}$
- Observation space Y
- State equations

$$\begin{cases} x(k+1) = F(x(k), u(k)) \\ x(0) = x_0 \\ y(k) = G(x(k), u(k)) \end{cases}$$

Remarks

- state definition must simplify control design
- Unknown quantities can be learned with input-output measurements

Definition

- **X**,**U**,**Y** usually subsets of euclidean spaces
- State equations

$$\begin{cases} x(k+1) = Fx(k) + Gu(k) \\ x(0) = x_0 \\ y(k) = Hx(k) + Eu(k) \end{cases}$$

Linear State feedback

$$u(k) = -Kx(k), \text{ K : gain matrix}$$

Objectives

Evaluate

- long-run behavior of state process
- metric performance (total cost)

Plan

- definitions (total cost, equilibria)
- Stability results (discrete, continuous case)
- Application (linear case)

Total cost

$$J(x) = \sum_{k=0}^{\infty} c(x(k)), x(0) = x \in \mathbf{X}$$

- If J is finite, stability typically guaranteed
- related to average cost (optimal control)
- ”forward looking”

Definition : stable in the sense of Lyapunov

The equilibrium x^ε is stable in the sense of Lyapunov if for all ε, there exists $\delta > 0$ such that if $\|x_0 - x^\varepsilon\| < \delta$, then

$$\|\mathcal{X}(k, x_0) - \mathcal{X}(k, x^\varepsilon)\| < \varepsilon, \text{ for all } k \geq 0$$

Asymptotic Stability

An equilibrium x^ε is said to be asaymptotically stable if x^ε is stable in the sense of Lyapunov and for some $\delta_0 > 0$, whenever $\|x_0 - x^\varepsilon\| < \delta_0$

$$\lim_{k \to \infty} \mathcal{X}(k, x_0) = x^\varepsilon$$

*The set of x_0 for which the limit holds : **region of attraction**.*
*The equilibrium is **globally asymptotically** stable if the region of attraction is* **X**.

Definition

- non-negative
- decreasing (drift inequality)
- Frequently
  > inf-compact : $\{x \in \mathbf{X} : V(x) \leq V(x^0)\} \subset \mathbf{X}$ bounded $\forall x^0 \in \mathbf{X}$
  > coercive : $\lim_{\|x\| \to \infty} V(x) = \infty$

Total cost is Lyapunov under mild conditions

*Suppose c and J are **non-negative** and **finite valued**, then*

- $J(x(k))$ *is non-increasing and* $\lim_{k \to \infty} J(x(k)) = 0, \forall x_0$,
- *If J is also* **continuous**, **inf-compact**, *and* **vanishes** *only at* x^ε, *then* $\forall x, \lim_{k \to \infty} x(k) = x^\varepsilon$

Drift inequality considered

Poisson's inequality

$$V(F(x)) \leq V(x) - c(x) + \hat{\eta}, \hat{\eta} \geq 0$$

Proposition 2.4

Suppose the inequality holds for $\hat{\eta} = 0$, V is ***continuous****,* ***inf-compact****, with a* ***unique minimum*** *at x^{ε}. Then, the equilibrium is stable.*

Comparison theorem (2.5)

Poisson's inequality implies the following bounds

- *For each* $N \geq 1$ *and* $x = x(0)$

$$V(x(N)) + \sum_{k=0}^{N-1} c(x(k)) \leq V(x) + N\hat{\eta}$$

- *If* $\hat{\eta} = 0$*, then* $J(x) \leq V(x), \forall x$
- *Suppose that* $\hat{\eta} = 0$*,* V*,* c *are* ***continuous****. Suppose also* c *is* ***inf-compact****,* ***vanishes only*** *at* x^{ε}*. Then the equilibrium is globally asymptotically stable.*

Can we reach equality ?

Yes ! Proposition 2.6: If

- $V(F(x)) = V(x) - c(x)$
- J is **continuous**, **inf-compact**, and vanishes **only at** x^ε
- V is continuous

Then $J(x) = V(x) - V(x^\varepsilon)$

Assumptions

- $x(k+1) = Fx(k)$
- $c(x) = x^T S x, S \in \mathcal{S}^+(\mathbb{R}^n)$

Lyapunov equation

$$M = S + F^T M F$$

Proposition 2.10

The following are equivalent

- The origin is locally asymptotically stable
- The origin is globally asymptotically stable
- The Lyapunov equation admits a solution $M \geq 0$ for any $S \geq 0$
- Each eigenvalue λ of F satisfies $|\lambda| < 1$

First summary

We have seen

- the **State space model** (observations from inputs)

$$\begin{cases} x(k+1) = Fx(k) + Gu(k) \\ x(0) = x_0 \\ y(k) = Hx(k) + Eu(k) \end{cases}$$

- In some case, explicit **policy** (inputs from observations)

$$u(k) = -Kx(k), \text{ K : gain matrix}$$

- Tools for assessing stability (cf. total cost and Lyapunov functions)
  - cf :*Each eigenvalue λ of $F - GK$ satisfies* $|\lambda| < 1$

Actors and critics

- actors : $\{\phi^\theta : \theta \in \mathbb{R}^d\}$
- critics computes exactly J_θ
- actor-critic algorithm

$$\theta^* = \arg\min_\theta < v, J_\theta >$$

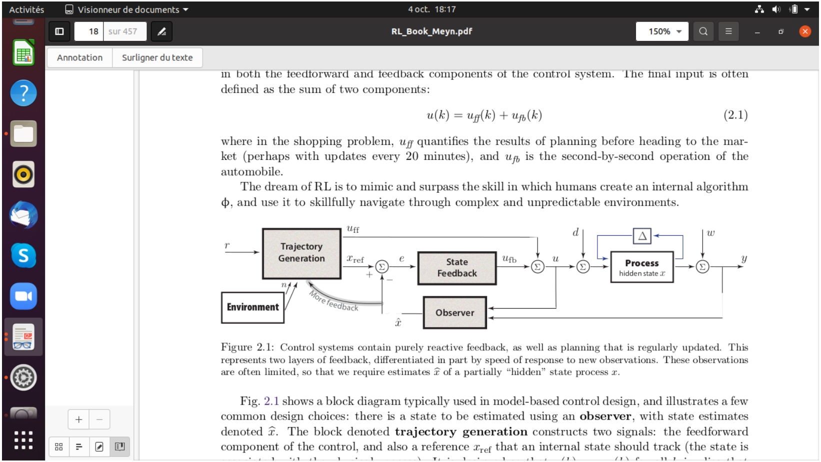
Temporal differences

How can we estimate a value function without a model ?

$$\mathcal{D}_{k+1}(\hat{J}) := -\hat{J}(x(k)) + \hat{J}(x(k+1)) + c(x(k), u(k)), k \geq 0,$$

with $u(k) = \phi^\theta(x(k))$

How ignoring noise ?



in both the feedforward and feedback components of the control system. The final input is often defined as the sum of two components:

$$u(k) = u_{ff}(k) + u_{fb}(k) \tag{2.1}$$

where in the shopping problem, u_{ff} quantifies the results of planning before heading to the market (perhaps with updates every 20 minutes), and u_{fb} is the second-by-second operation of the automobile.

The dream of RL is to mimic and surpass the skill in which humans create an internal algorithm ϕ, and use it to skillfully navigate through complex and unpredictable environments.

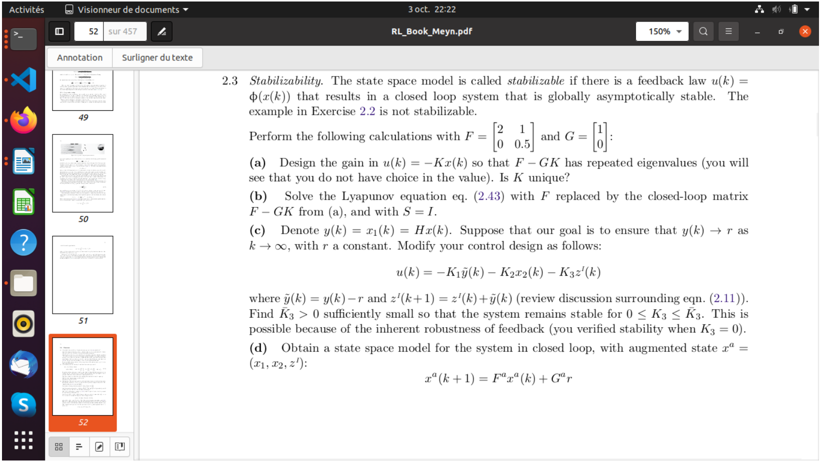
Figure 2.1: Control systems contain purely reactive feedback, as well as planning that is regularly updated. This represents two layers of feedback, differentiated in part by speed of response to new observations. These observations are often limited, so that we require estimates $\hat{x}$ of a partially "hidden" state process x.

Fig. 2.1 shows a block diagram typically used in model-based control design, and illustrates a few common design choices: there is a state to be estimated using an **observer**, with state estimates denoted $\hat{x}$. The block denoted **trajectory generation** constructs two signals: the feedforward component of the control, and also a reference x_{ref} that an internal state should track (the state is

03

Exercise 2.3

One example : exercise 2.3



2.3 *Stabilizability.* The state space model is called *stabilizable* if there is a feedback law $u(k) = \phi(x(k))$ that results in a closed loop system that is globally asymptotically stable. The example in Exercise 2.2 is not stabilizable.

Perform the following calculations with $F = \begin{bmatrix} 2 & 1 \\ 0 & 0.5 \end{bmatrix}$ and $G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$:

(a) Design the gain in $u(k) = -Kx(k)$ so that $F - GK$ has repeated eigenvalues (you will see that you do not have choice in the value). Is K unique?

(b) Solve the Lyapunov equation eq. (2.43) with F replaced by the closed-loop matrix $F - GK$ from (a), and with $S = I$.

(c) Denote $y(k) = x_1(k) = Hx(k)$. Suppose that our goal is to ensure that $y(k) \to r$ as $k \to \infty$, with r a constant. Modify your control design as follows:

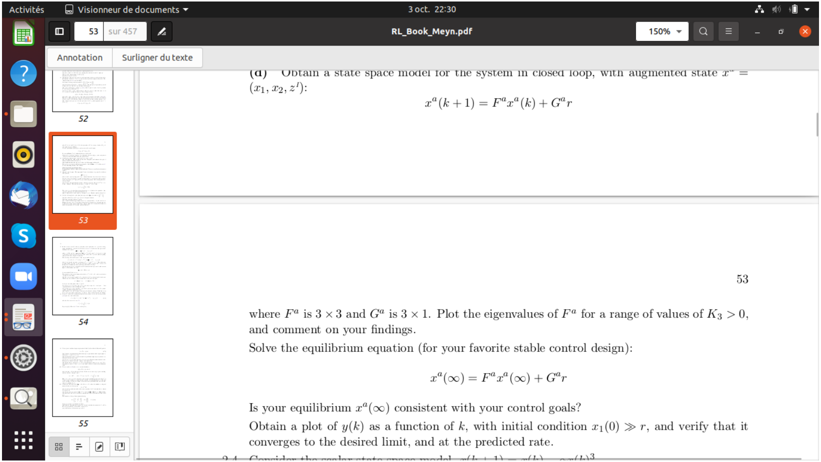
$$u(k) = -K_1\tilde{y}(k) - K_2 x_2(k) - K_3 z^I(k)$$

where $\tilde{y}(k) = y(k) - r$ and $z^I(k+1) = z^I(k) + \tilde{y}(k)$ (review discussion surrounding eqn. (2.11)). Find $\bar{K}_3 > 0$ sufficiently small so that the system remains stable for $0 \leq K_3 \leq \bar{K}_3$. This is possible because of the inherent robustness of feedback (you verified stability when $K_3 = 0$).

(d) Obtain a state space model for the system in closed loop, with augmented state $x^a = (x_1, x_2, z^I)$:

$$x^a(k+1) = F^a x^a(k) + G^a r$$

One example : exercise 2.3



(d) Obtain a state space model for the system in closed loop, with augmented state $x^a =$ (x_1, x_2, z^I):

$$x^a(k+1) = F^a x^a(k) + G^a r$$

53

where F^a is 3×3 and G^a is 3×1. Plot the eigenvalues of F^a for a range of values of $K_3 > 0$, and comment on your findings.

Solve the equilibrium equation (for your favorite stable control design):

$$x^a(\infty) = F^a x^a(\infty) + G^a r$$

Is your equilibrium $x^a(\infty)$ consistent with your control goals?

Obtain a plot of $y(k)$ as a function of k, with initial condition $x_1(0) \gg r$, and verify that it converges to the desired limit, and at the predicted rate.

2.4 Consider the scalar state space model $x(k+1) = x(k) - \alpha x(k)^3$

roots behaviour

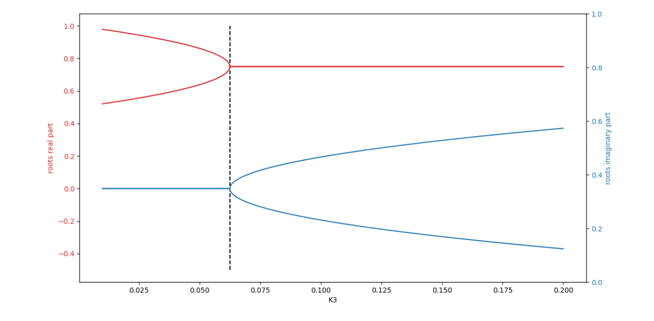


Figure: roots of $\lambda^2 - \frac{3}{2}\lambda + 0.5 - K_3, K_3 \in [0.01, 0.2]$

convergence simulations with
$K_3 = 0.01, K_2 = 1, r = 0.5, x_1^0 = 20, x_2^0 = 2, x_3^0 = 4$

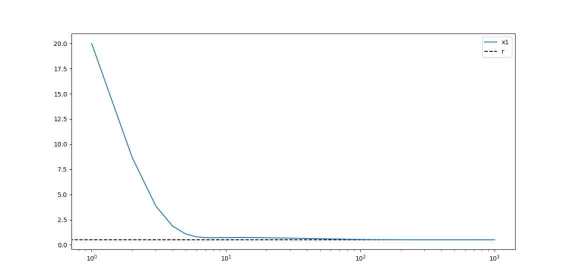


Figure: k$\rightarrow x_1^k, k \in \{1, ..., 1000\}$

- the **State space model** (observations from inputs)

$$\begin{cases} x(k+1) = Fx(k) + Gu(k) \\ x(0) = x_0 \\ y(k) = Hx(k) + Eu(k) \end{cases}$$

- In some case, explicit **policy** (inputs from observations)

$$u(k) = -Kx(k), \text{ K : gain matrix}$$

- Tools for assessing stability (cf. total cost and Lyapunov functions)
  - > cf :*Each eigenvalue* λ *of* $F - GK$ *satisfies* $|\lambda| < 1$

From Control to RL

- Actor-critic
- TD Learning
- Exploration and exploitation paradigm

Taking into account of noise

- architecture not sensitive (wrt disturbance class)
- assumptions (frequency domains, Lyapunov drift inequalities etc.)