

# Control prin învățare (Learning Control)

## Laborator 4: Algoritmi online cu aproximare

**Regulament** – același ca și la laboratorul precedent, cu excepția termenului limită pentru predarea soluției, care este **27 mai 2018**. Acesta este și **termenul limită pentru orice soluție întârziată**. Studenții care la această dată nu au toate soluțiile predate nu vor fi eligibili pentru examen.

### Introducere

Vom considera din nou problema pendulului inversat, vezi laboratorul 3 pentru o descriere. Codul ce implementează simulatorul pendulului inversat și formează baza laboratorului poate fi descărcat de la adresa:

<http://busoniu.net/teaching/ci2018>

secțiunea “Practical assignments”. Dezarhivați codul într-un director de pe calculatorul dvs, navigați din MATLAB în acest director și rulați scriptul `startup`. Implementarea algoritmilor de planificare online este deja furnizată, iar fișierul `lab4_example` ilustrează cum se pot rula acești algoritmi și cum se pot schimba parametrii lor.

Sistemul este vizualizat în timp ce algoritmi rulează, pentru a avea o impresie în “timp real” a performanței lor, dar această vizualizare poate fi dezactivată dacă se dorește ca algoritmi să ruleze mai repede. După ce un algoritm finalizează un experiment, acesta este reprezentat grafic, oferind o imagine de ansamblu a performanței. Fișierul `exemplu` arată și cum se poate afla returnul obținut de-a lungul traiectoriei și timpul mediu de execuție al planificării (media este efectuată de-a lungul pașilor traiectoriei).

### Cerințe

Vom investiga întâi comportamentul algoritmului de *învățare  $Q$  aproximată*. Fiindcă algoritmul depinde de explorare, care este aleatoare, analiza comportamentului său va fi una calitativă.<sup>1</sup> Începeți prin a rula algoritmul cu setările furnizate.

1. Studiați calitativ efectul ratei de învățare  $\alpha$ , folosind o valoare mică, o valoare aproape sau egală cu 1, și o valoare intermediară (toate ținute constante de-a lungul învățării). **[2p]**
2. Studiați calitativ efectul ratei de scădere a eligibilității  $\lambda$ , folosind valoarea 0, o valoare aproape de 1, și o valoare intermediară. **[2p]**
3. Studiați calitativ efectul ratei de scădere a explorării `explordecay` (probabilitatea de explorare scade la sfârșitul fiecărei traiectorii cu o rată dată de această valoare). **[2p]**

Vom investiga și algoritmul *LSPI online*.

4. Studiați calitativ efectul intervalului  $K$  între îmbunătățirile legii de control, folosind de exemplu 10, 100, 1000 pași (de notat că în cod se setează parametrul echivalent după, care este egal cu numărul de pași  $K$  înmulțit cu timpul de eșantionare). Puteți de exemplu să schimbați și secvența de explorare, controlând `explordecay`. **[2p]**

---

<sup>1</sup>O analiză cantitativă ar necesita rularea fiecărui experiment de un număr mare de ori pentru a obține rezultate semnificative din punct de vedere statistic, dar aceasta nu se încadrează în aria acestui laborator.

Pentru un nivel de încredere mai bun în rezultate, fiecare experiment poate fi repetat de câteva ori (rezultatele depinzând de explorare și fiind așadar aleatoare). Pentru aceste experimente multiple, vi se recomandă dezactivarea vizualizării conform instrucțiunilor de mai sus, pentru a obține rezultatele mai rapid.

Pentru ambii algoritmi, încercați să obțineți o performanță mai bună decât cea obținută cu valorile furnizate ale parametrilor. Discutați procesul de acordare, rezultatele obținute, precum și dificultățile întâmpinate.