Content

- Algorithms testing
  - Random walk (classic)
  - Random walk (modified)
  - Cliff
- Roadmap: where are we going?
- Future lines (18/05/2017)

Content

- **ALGORITHMS TESTING**
  - Random walk (classic)
  - **RANDOM WALK (MODIFIED)**
  - **CLIFF**
- Roadmap: where are we going?
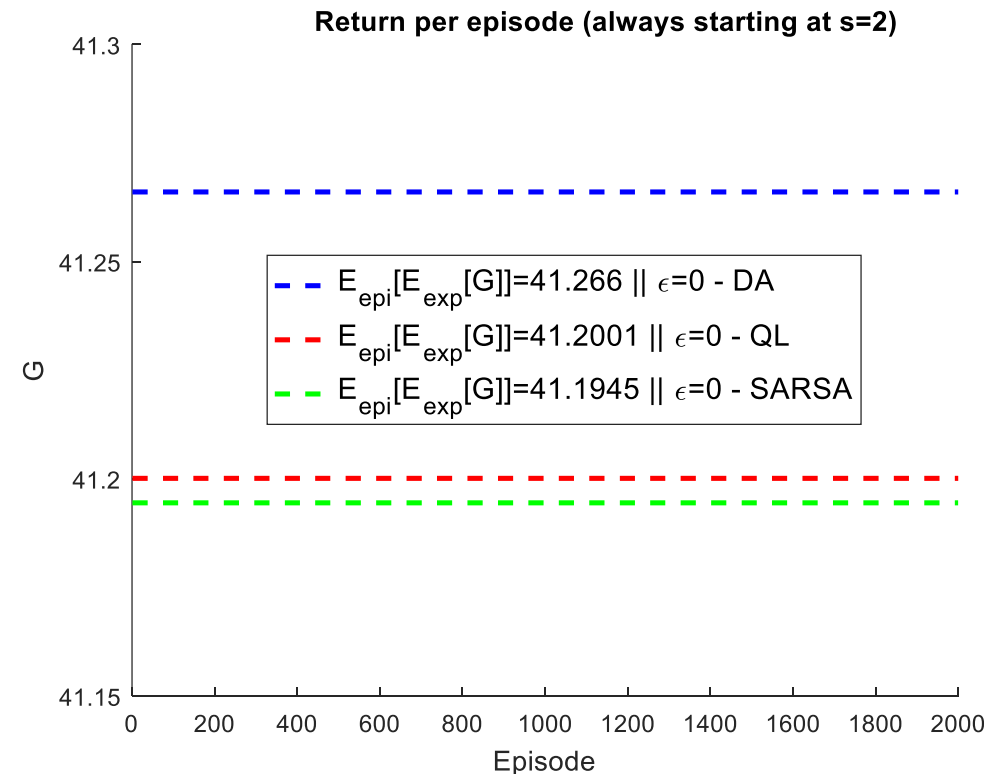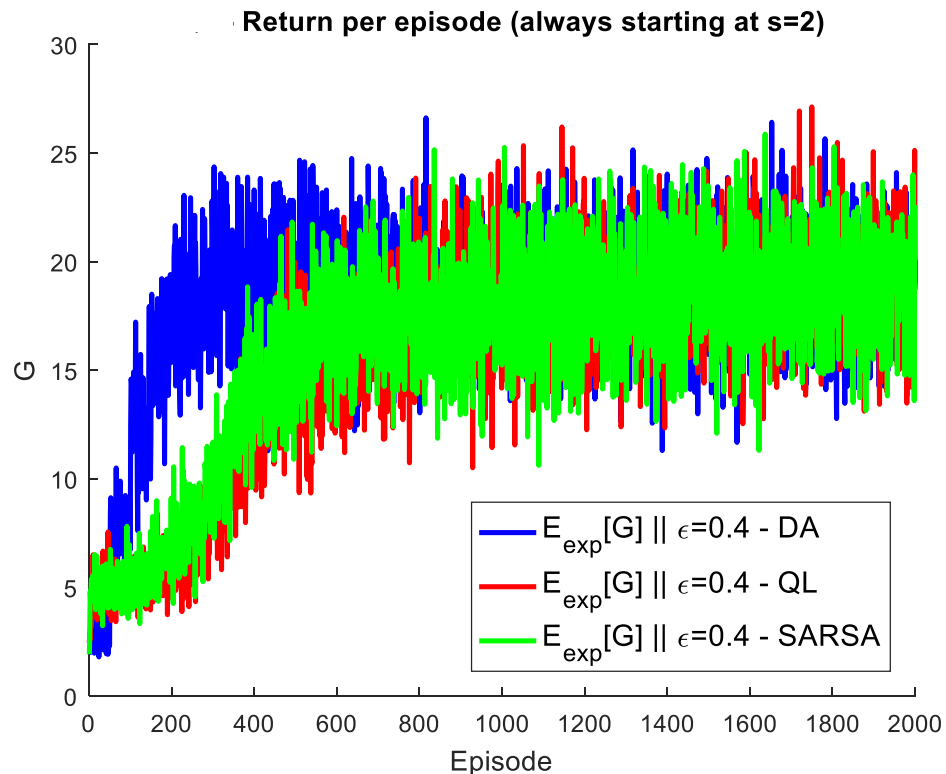- Future lines (18/05/2017)

Random walk (modified)

- Modified version
  - Nstates/2 states with random reward between 1 and 2. Last state on the right more reward, 150, to assure convergence.
  - Nstates parametrizable (**Nstates = 13** was chosen for testing)
  - Always starting at state 2 (first state on the left)
  - Optimal policy: finishing at T2
  - Random transition matrix (i.e: when action is "left", you go **left** with $p_{left} = 0.8 - 1$ probability and **right** with $p_{right} = 1 - p_{left}$ probability)

Comparison

- Comparison between DA TD (model-free), SARSA and Q-learning. All of them with optimal parameters.
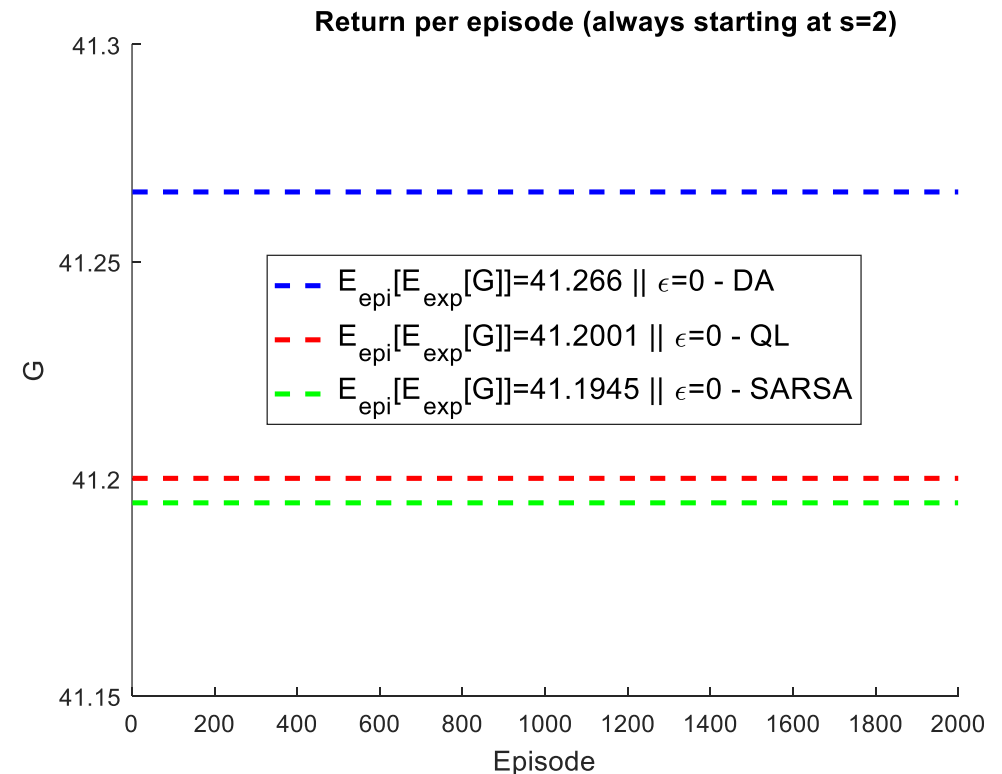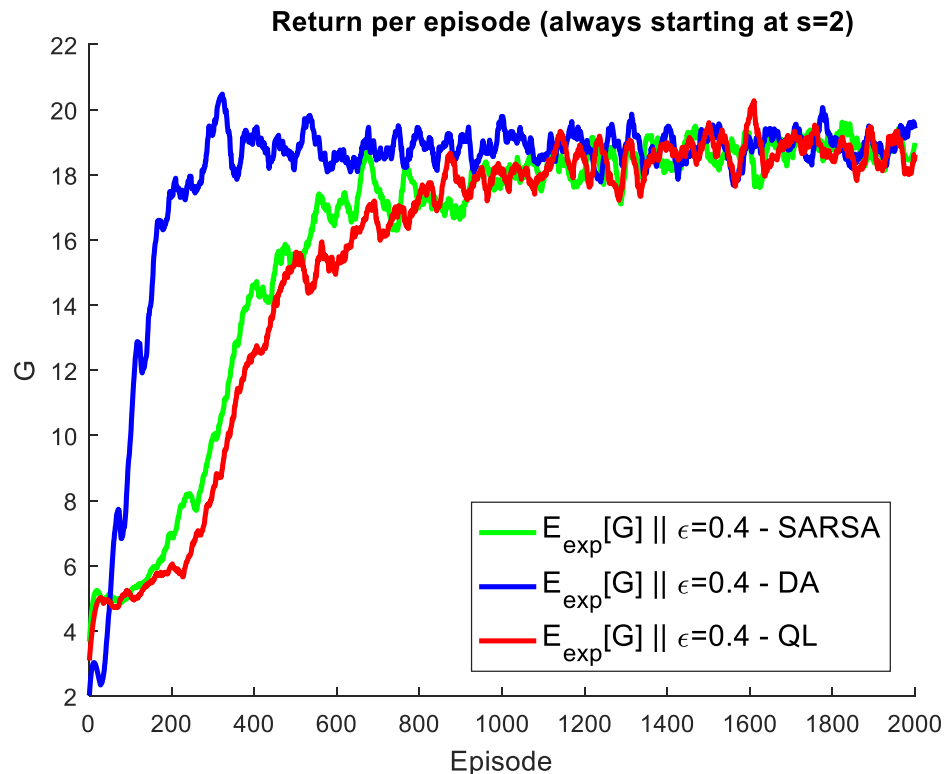
¿Comparar también con model-based?

Return per episode (always starting at s=2)

$E_{exp}[G] \| \epsilon=0.4$ - DA
$E_{exp}[G] \| \epsilon=0.4$ - QL
$E_{exp}[G] \| \epsilon=0.4$ - SARSA

Return per episode (always starting at s=2)

$E_{epi}[E_{exp}[G]]=41.266 \| \epsilon=0$ - DA
$E_{epi}[E_{exp}[G]]=41.2001 \| \epsilon=0$ - QL
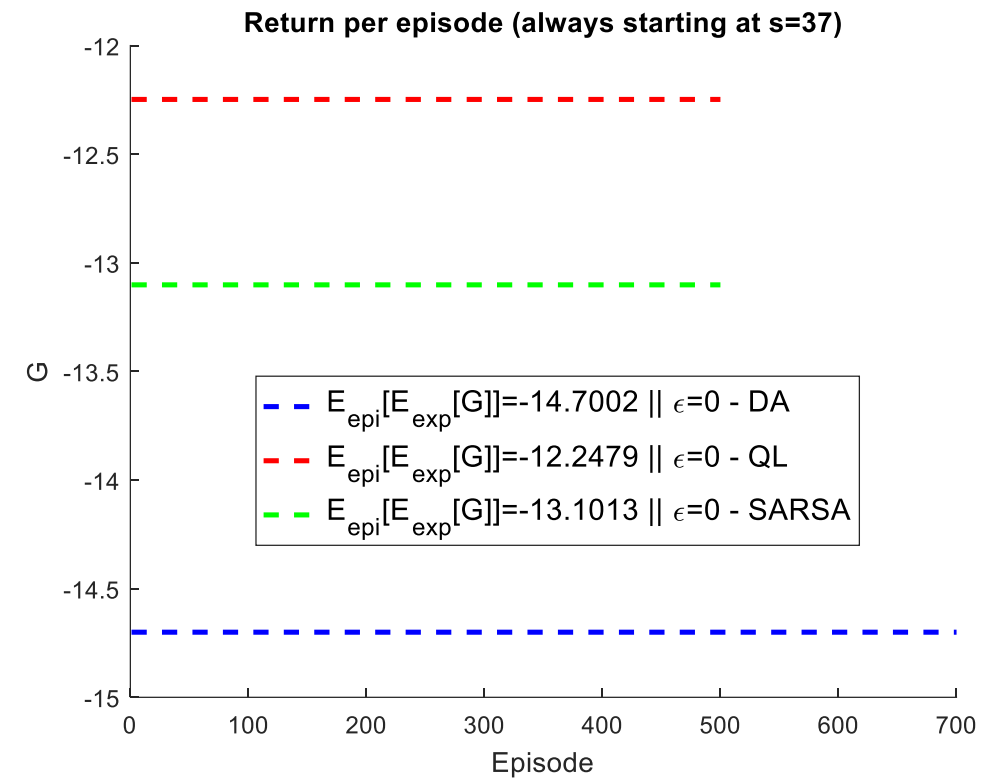$E_{epi}[E_{exp}[G]]=41.1945 \| \epsilon=0$ - SARSA

Comparison

- Comparison between DA TD (model-free), SARSA and Q-learning. All of them with optimal parameters.
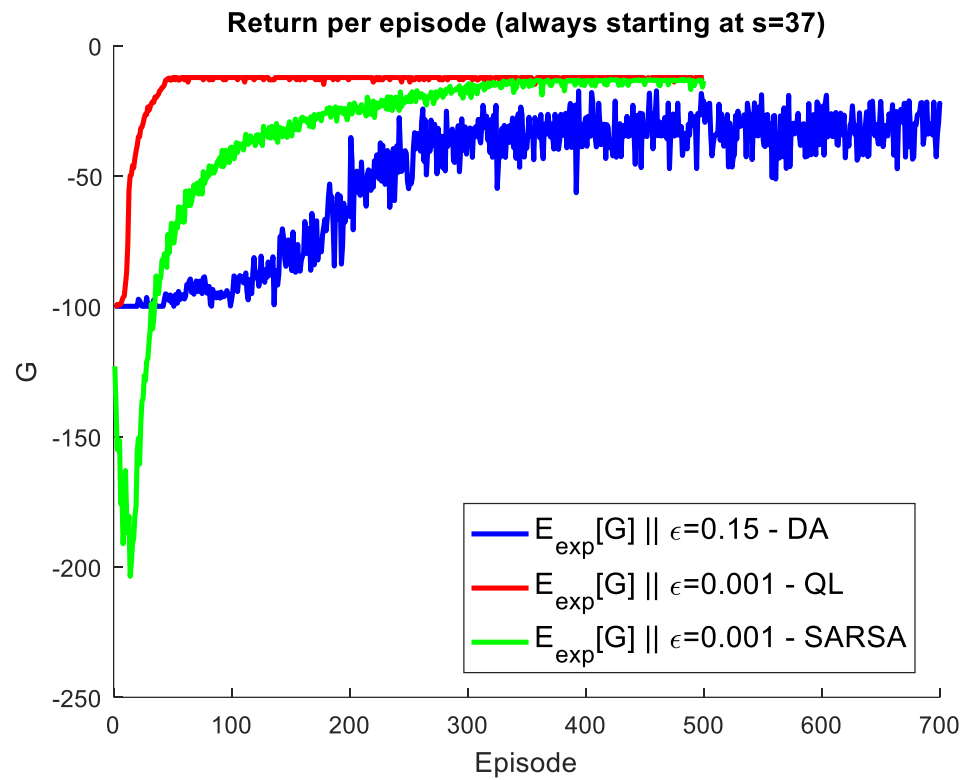
¿Comparar también con model-based?

Return per episode (always starting at s=2)

$E_{exp}[G] \parallel \epsilon=0.4$ - SARSA
$E_{exp}[G] \parallel \epsilon=0.4$ - DA
$E_{exp}[G] \parallel \epsilon=0.4$ - QL

Return per episode (always starting at s=2)

$E_{epi}[E_{exp}[G]]=41.266 \parallel \epsilon=0$ - DA
$E_{epi}[E_{exp}[G]]=41.2001 \parallel \epsilon=0$ - QL
$E_{epi}[E_{exp}[G]]=41.1945 \parallel \epsilon=0$ - SARSA

Cliff

Content

- Algorithms testing
  - Random walk (classic)
  - Random walk (modified)
  - Cliff
- **ROADMAP: WHERE ARE WE GOING?**
- Future lines (18/05/2017)

Roadmap: where are we going?

Algorithm…	Compare with…	Problems to test with…

1. Dual-Ascent (DA):
   1. Model based → exact solution (DP)
   2. Model free → stochastic gradient (RL)
	▪ SARSA
	▪ Q-learning
	▪ Double Q-learning
		▪ Random walk (classical)
		▪ Random walk (modified)
		▪ Cliff

2. DA with Linear Function Approximation (DA-LFA)	▪ LSPI ▪ LSTD ▪ GTD2 ▪ Variance reduction	▪ Chain walk (RL course version) ▪ Mountain car → *¡¡CONTINUOUS STATE SPACE!!*
3. DA with Non-Linear Function Approximation (DA-NLFA)	▪ Deep Q-learning (DQN) ▪ Neural fitted Q-iteration (NFQ)	▪ Mountain Car ▪ Cart-Pole ▪ Videogames Neural networks / Python / OpenAI

Content

- Algorithms testing
    - Random walk (classic)
    - Random walk (modified)
    - Cliff
- Roadmap: where are we going?
- **FUTURE LINES (18/05/2017)**

Future lines (18/05/2017)

- Complete simulations