OLCAR - Exercise 3 – Reinforcement Learning

Answers to question related to programming exercise

Handout: 12.05.2015

Due: 26.05.2015

Yi Hao Ng, Raphael Stadler

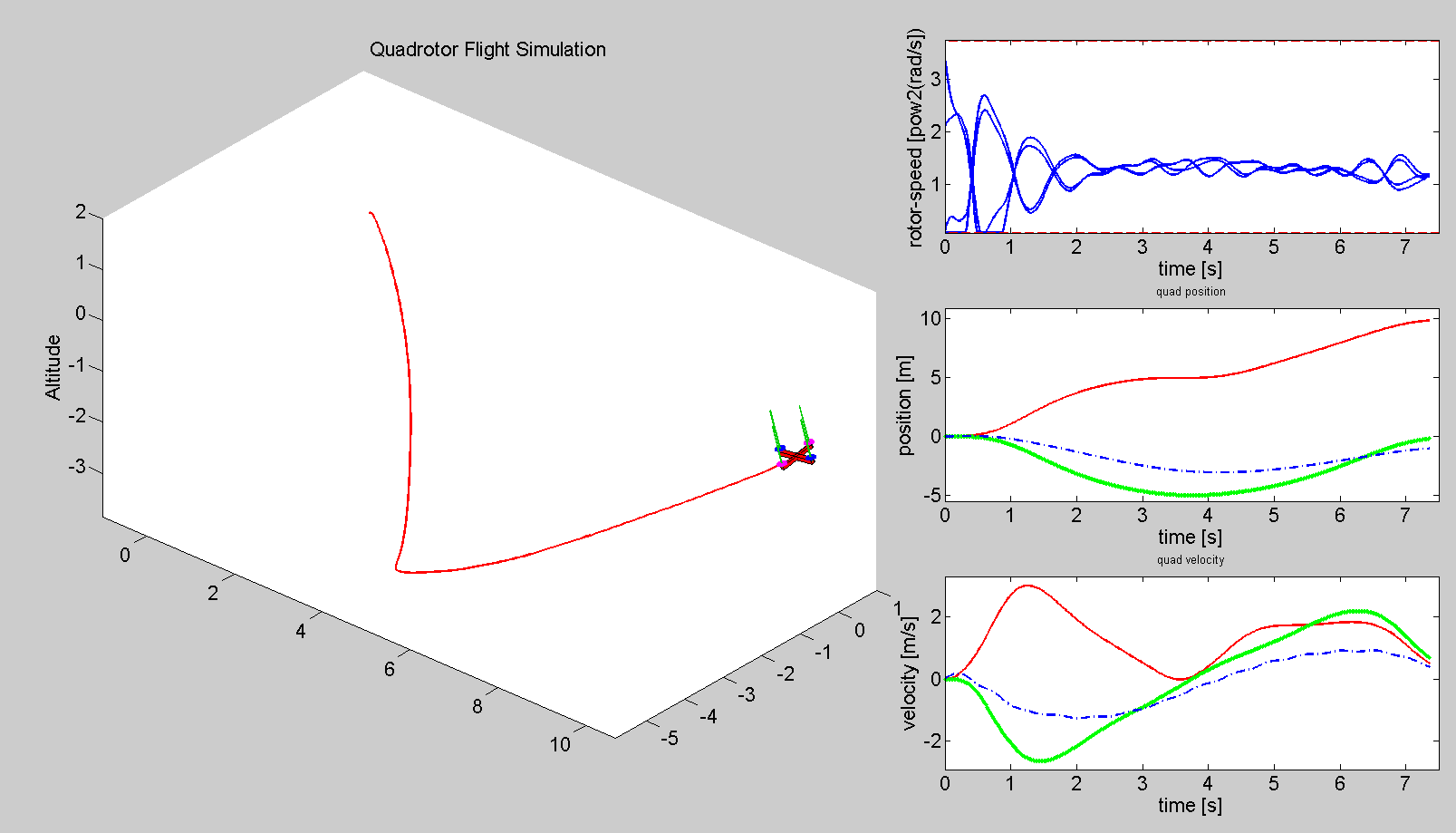
(Team Number 10)

1. How much cost improvement did you obtain using PI2 learning?

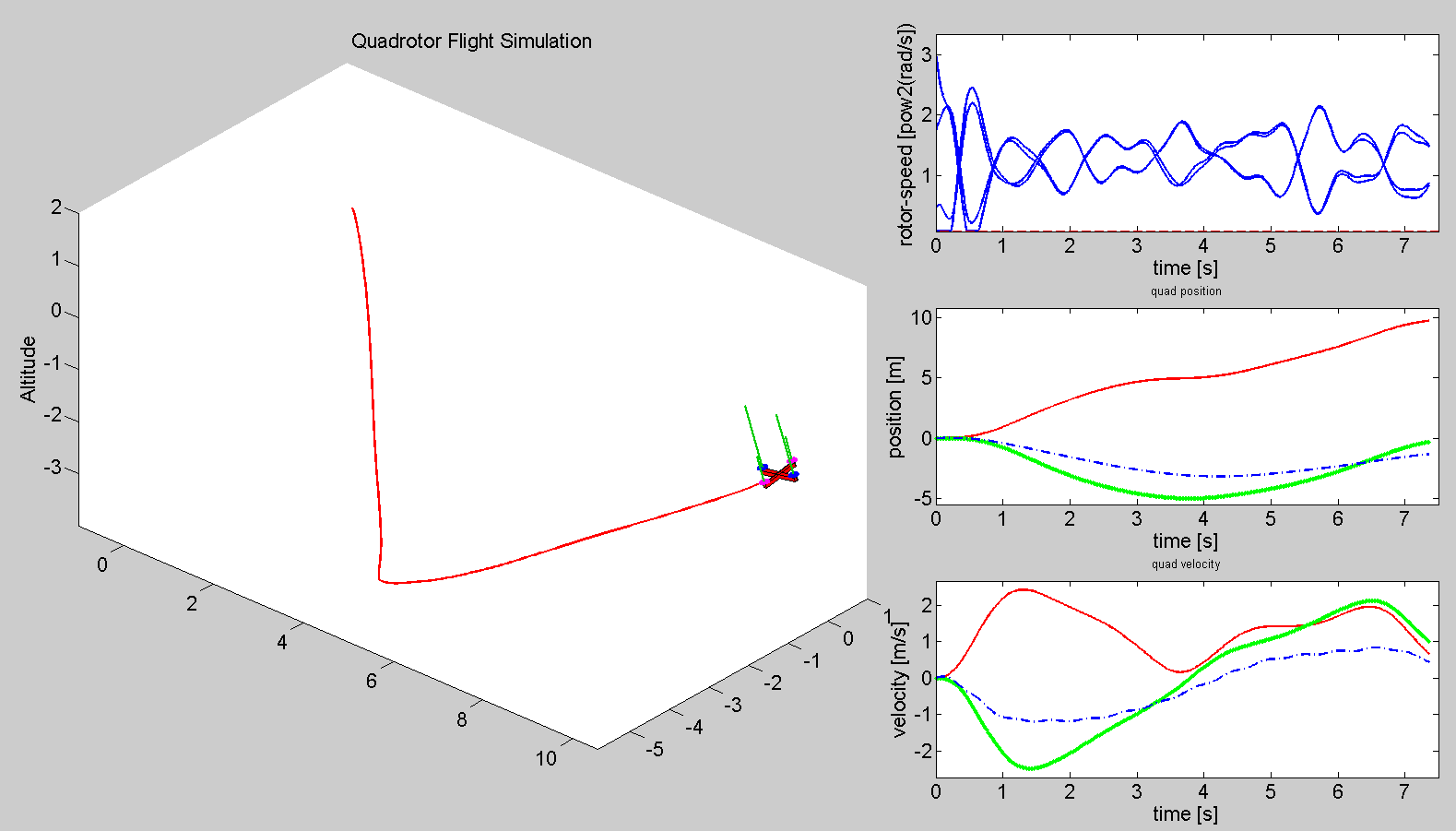
(answer in 1 sentence and attach one of your cost-plots)

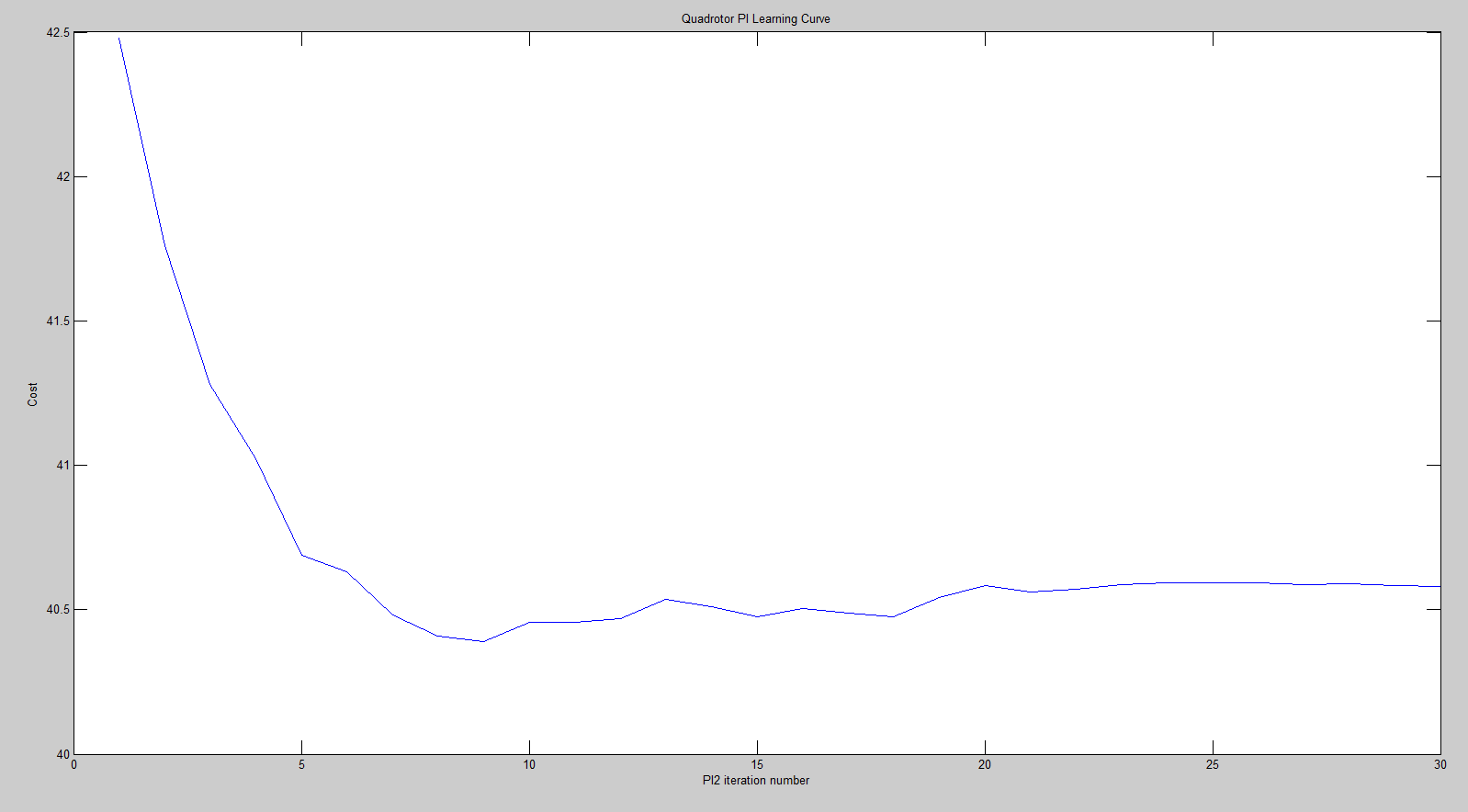
**Currently: Reduced cost by 5 % (2 out of 40).**

Task.std\_noise = 0.0015; % standard deviation of exploration noise for PI2

ILQC: 

PI2:





2. How does the exploration noise (Task.std\_noise) affect the learning curve? What happens if you decrease/increase it?

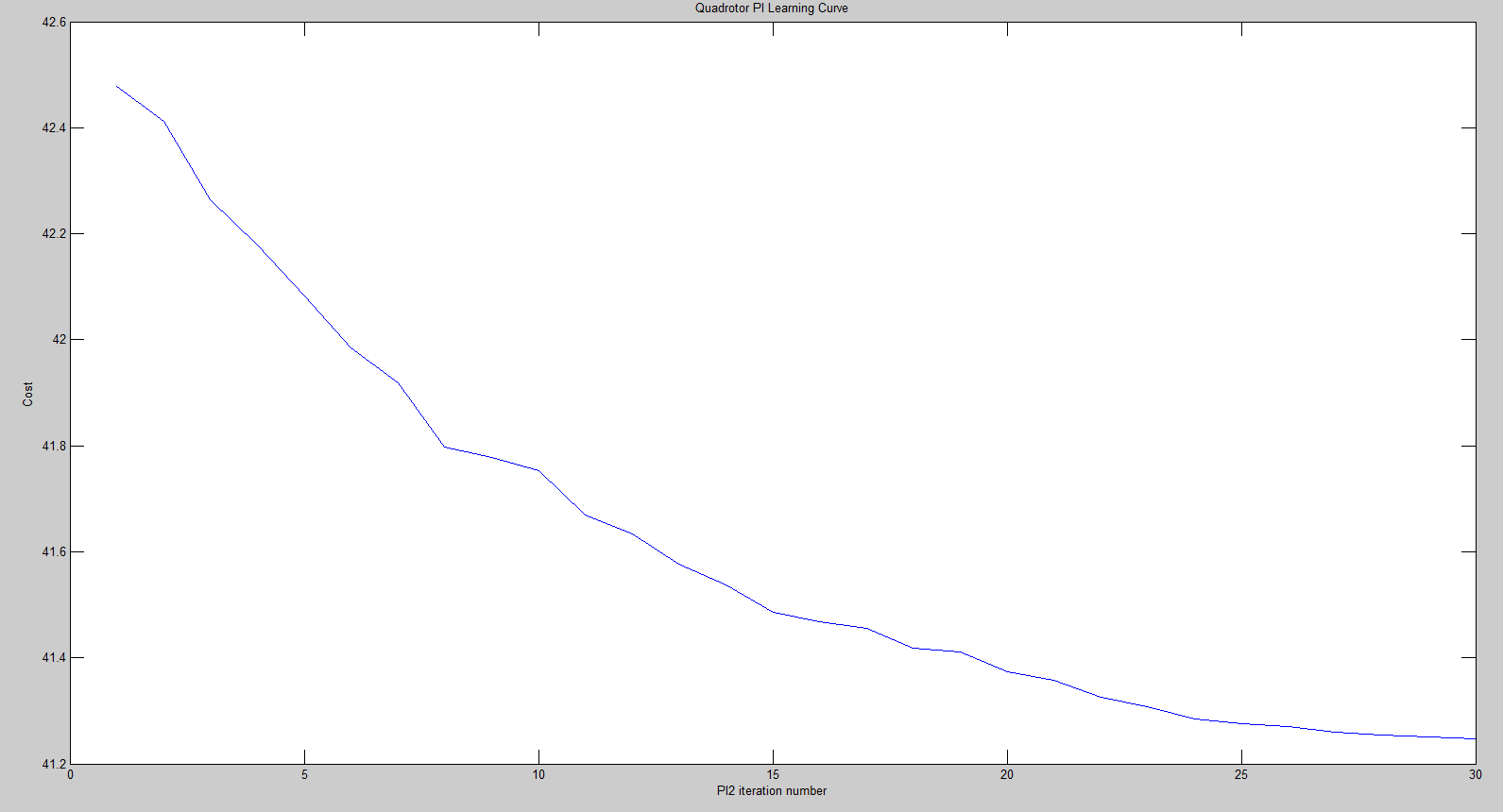
Task.std\_noise = 0.0015; % standard deviation of exploration noise for PI2

Increasing: ode45 crashs!

* Bigger “step sizes” are possible.

Decreasing: slows down convergence:

Task.std\_noise = 0.00015; % standard deviation of exploration noise for PI2



3. The tuning parameter Task.num\_reuse specifies how many (of the best) rollouts are saved, carried over and reused in the next learning iteration. Why does it make sense to keep some of the best rollouts for the next update?

Instead of greedily optimize, you can do better by taking a history into account in such a way.

4. How does the quality of your initial guess affect the PI2 learning? For example, what happens if you limit your ILQC iterations to only 1?

* Initial guess has direct effect on the quality.

5. While executing your program, you might have noticed that the cost is not always strictly decreasing during learning. What is your explanation for this behaviour?

* Exploration: perturbed epsilon