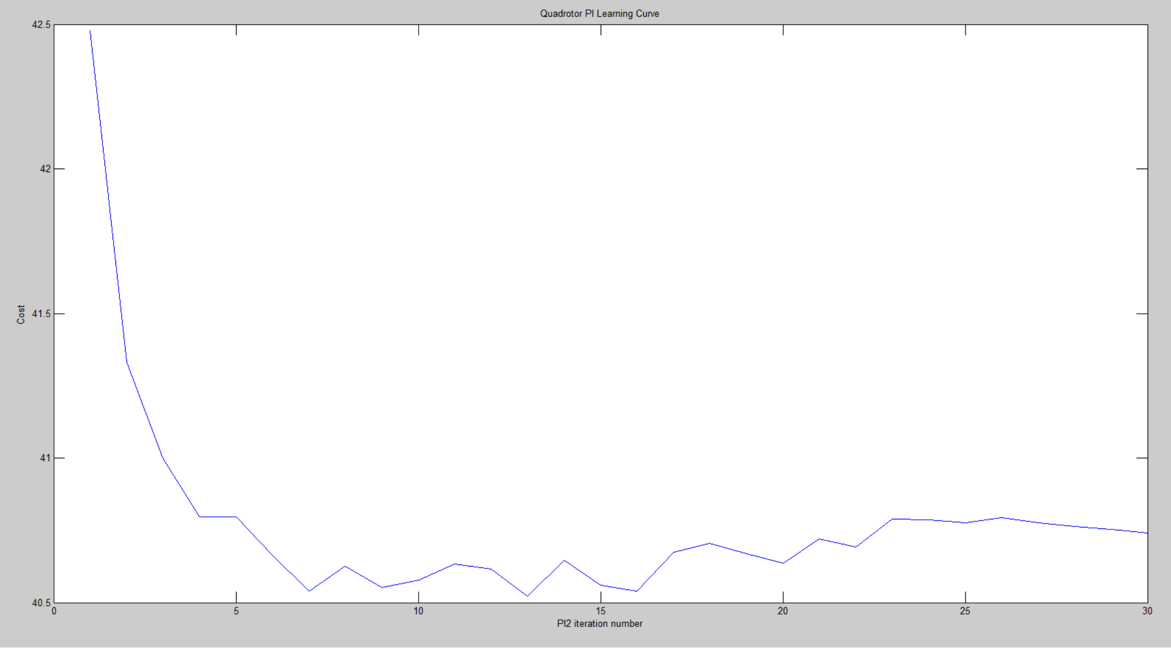
OLCAR- Exercise 3 – Reinforcement Learning

Answers to question related to programming exercise

Handout: 12.05.2015	Yi Hao Ng, Raphael Stadler
Due: 26.05.2015	(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)

By using PI2 the cost could be reduced by approximately 5 % (from 42.5 to 40.5).



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

If the noise is chosen to be too big (e.g. to 0.015) at some part of the algorithm the integration using ode45 fails. Decreasing the exploration noise (e.g. to 0.00015) makes the algorithm converge slower.

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?

If the algorithm succeeded to find a good solution with a specific set of basis functions and its corresponding parameters, this solution can be taken as a basis to do further optimizations. In the end, the goal of the algorithm is to find a global solution to the cost minimization problem. Since the costs do not strictly decrease, it gets meaningful to also take the minimal cost (of the last rollouts) into account instead of simply taking the last cost.

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?

The performance of the PI2 learning is directly affected by the quality of the initial guess. If the initial guess is of bad quality, the cost of the initial guess is really big (as it is e.g. the case when only 1 ILQC iteration is performed, which even makes the quadrotor “crash”). Having such a big initial value causes the algorithm not to be able to converge (in the allowed maximum number of iterations) and the cost oscillates a lot throughout the iterations.

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?

During the PI2 algorithm, when choosing the delta theta (the difference of the basis function parameters) the probabilistic exploration noise is also included in the calculations. Because of this stochasticity, exploration gets possible. For such an exploration it is possible that the found solution is not as good as the one from the previous iteration.