**Reinforcement Learning**

**RL algo: Deep deterministic policy gradient (All simulations till 10 seconds)**

**Train set:** 100 randomly generated initial positions. (100 episodes)

For the plot below:

Yellow line: expected Q value

Dark blue line: Average reward for last 100 episodes

Light blue line: Episode reward.

Chart

Description automatically generated

Observation: The average reward increases, and Q value seems to be converging to the true value. Needs to train on more episodes.

**Test set**: 50 randomly generated initial positions (with a different seed from train set)

ISE on test set for the 3 control methods

|  |  |  |  |
| --- | --- | --- | --- |
|  | **LQR** | **Bryson\_LQR** | **RL DDPG** |
| x | 7711.9 | 5645.5 | 2765.3 |
| y | 8067.1 | 5970.4 | 2756.6 |
| z | 7314.1 | 4985.9 | 2285.6 |
| phi | 5.2 | 14.5 | 624.3 |

Observation: RL DDPG reduces the ISE considerably. The ISE for psi increases as we have not normalized the data before training or calculating the reward. Next step can be to normalize the data and train again.

Sample Plot: Generated by first set of initial conditions in the test set.

Chart

Description automatically generated with medium confidence