* Show that objective with max[KL(pi\_old, pi)] forms a lower bound on the performance of the policy [2]
* **Monotonic** improvement of TRPO? [2]
* Show that constant beta in the objective function does not suffice with first order methods [2]
* Without a constraint, maximization of L CP I would lead to an excessively large policy update [3]
* we can use a truncated version of generalized advantage estimation: Demo with TD error replacement [5]
* Insights about parallelization of the algorithm
* OpenAI Gym? [6]

Experiments:

6.1 Comparison of Surrogate Objectives First, we compare several different surrogate objectives under different hyperparameters. Here, we compare the surrogate objective L CLIP to several natural variations and ablated versions.

- No clipping or penalty

- Clipping

- KL penalty (fixed or adaptive)

6.2 Comparison of PPO with other algorithms in the continuous domain.