

Extended Abstract

CS294-112

Brian Cera

Edward Zhu

Learning Locomotion Policies for Tensegrity Robots

Tensegrity robots are a novel class of robots that, while possessing certain mechanically advantageous properties, present a challenging problem in terms of robust control. Especially as there is growing interest in using such robots for locomotion, expressive end-to-end policies need to be developed for these high dimensional systems – policies that can handle rich sensor observation data and complex action inputs.

For our final project, we attempted to use model-based RL methods to learn locomotion policies for two different tensegrity robots. The first is a quadruped robot with an actuated, flexible tensegrity spine. The goal here was to achieve locomotion by leveraging the flexibility of the body despite using simple 1-degree-of-freedom legs. For the method, we used guided policy search which has shown success in generating periodic gaits for complex systems. Unfortunately, our attempt proved to be unsuccessful as the robot could not achieve stable locomotion after training, despite a decrease in training cost. Reflection and analysis on our method and results led us to believe that there were three main reasons for this failure: 1) The cost functions we tried were too simple and were not expressive enough to indicate locomotion success in the context of this robot. We observed that a simple cost function in this case would lead the robot to learn how to fall in the right direction and in some cases inch or crawl forward, but no meaningful policy emerged from training. 2) There was insufficient exploration of the state and action spaces due to the initial condition being an unstable equilibrium point. Because of this, most actions would lead to the robot falling down and states which were important to the success of locomotion were rarely reached. 3) The dynamics of the system are highly coupled and non-linear, with discontinuities caused by contact. This leads to poor local linear approximations which directly affect the supervision given to the global policy.

The second robot we investigated was a 6-bar spherical tensegrity robot. For this system, we used imitation learning with input remapping using an MPC expert. By formulating the locomotion problem as a constrained finite-time optimal control problem, we generated optimal trajectories which were used as supervision to train a neural network policy. Using this method, we were able to achieve stable locomotion with the learned policy which successfully emulated the MPC expert. In addition, we investigated the effects of randomizing system parameters during training on the robustness of policies. Specifically, we randomly sampled the parameter of rod mass and used that value to perform trajectory optimization, thus generating a set of trajectories optimized to succeed under various perturbed models. The policy trained with these samples achieved poorer performance on average, but appeared to have lower variance across a wider range of parameter values for mass. Due to time constraints, we were unable to match the number of samples used to train the original policy and attribute the poorer performance to the sparser samples, especially in the presence of an additional variable during the training process.