

Optimization-Based Policy Search for Control of Periodic Systems

Eric Whitman
Carnegie Mellon University
email: ewhitman@cmu.edu

SUMMARY

We present an optimization-based approach to control of periodic systems. We define an “energy” that is invariant along the desired trajectory. We then create a parametric policy function and use optimization to find good values for the parameters. We show results for two simulated systems: a hopping monopod and a simple model of a biped marching in place in the coronal plane. The resulting controllers perform well over a large region of the state space.

INTRODUCTION

Some dynamic tasks such as walking are periodic. For such periodic tasks, we often have a periodic trajectory that we wish the system to converge to regardless of the initial conditions. In order to properly handle startup transients and be robust to large perturbations, we also require a policy that is valid for a large region of the state space. If a quantity that is invariant over the desired trajectory can be defined, control can be looked at as regulation of this quantity.

METHODS

We start with a hopping monopod system modeled as a mass with a linear spring. The dynamics are given by:

$$\ddot{h} = \begin{cases} -mgh & ; h \geq 0 \\ -mgh + kh^2 + u; h < 0 \end{cases}.$$

The desired trajectory is hopping with a fixed energy where energy has the normal physical definition of kinetic energy plus gravitational potential energy plus spring potential energy. We distribute several control points along the desired trajectory and associate a gain k_i with each point. The policy (figure 1) is then given by $u = k_i(E - E_{des})$ where k_i is the gain associated with the nearest control point. The trajectory cost is defined as the weighted sum of squared control plus the sum of squared deviation in energy, and the objective function for a control policy is given by the sum of the costs of trajectories starting at several fixed locations. The k_i are then optimized using this objective function to produce a controller.

We also apply a similar technique to the control of a simple coronal plane biped marching in place. The biped legs are modeled as massless force and torque sources during both single and double support. During single support the biped acts as an inverted pendulum. The weight distribution between the two legs varies linearly with position during double support, resulting in continuous dynamics.

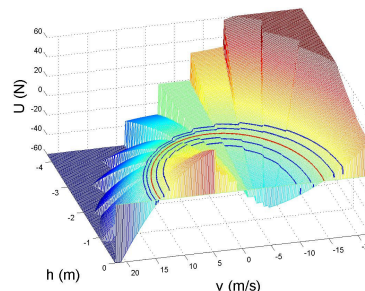


Figure 1: Hopper policy. The desired trajectory is shown in red and several controlled trajectories are shown in blue. Only the stance portion of the state space is shown.

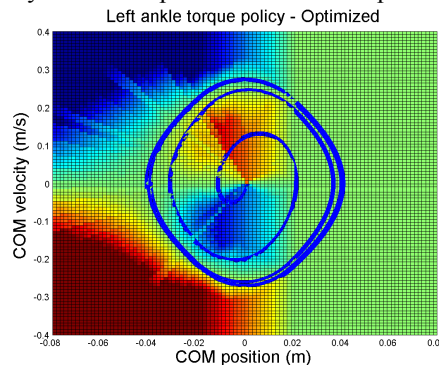


Figure 2: Left ankle policy for the coronal marching system. A trajectory starting from rest is shown.

We use a measure of “orbital energy” of the lateral position, y : $E = ay^2 + by^2$, which produces an elliptical desired trajectory. Each control point has two gains, one for each ankle. The optimized left ankle policy is shown in figure 2.

RESULTS AND DISCUSSION

For the hopping system, the energy controller actually outperforms a dynamic programming solution for trajectories that start very near the desired trajectory. This is possible because the energy controller is continuous, and unlike dynamic programming (DP), has no trouble with resolution. It performs well (costs within a few percent of a DP solution) for trajectories starting from between about 1/3 and twice the desired energy. Outside of this range, the performance degrades rapidly.

For the marching controller, the system performs well even well outside of the range in which it was trained. The policy was optimized using trajectories that started between 1/2 and 3/2 of the desired energy. The trajectory shown in blue, however, starts from rest (0 energy) and has a cost less than 10% higher than the trajectory generated by a DP policy.