Optimization-Based Policy Search for Control of Periodic Systems

Eric Whitman   Carnegie Mellon University

Introduction

Simple linear controllers often work well for regulation tasks (getting to and staying at a goal), but for periodic systems, the current target and the dynamics are typically time-varying. If we can define some quantity that is constant on the nominal trajectory (desired path), we can look at control of the periodic system as regulation of that quantity.

Parametric Optimization: A technique for picking the inputs of a function to minimize (or maximize) its output. In this case, we optimize the gains to produce trajectories with the lowest cost.

Dynamic Programming (DP): A non-parametric technique for coming up with control policies. It is computationally intensive, but very accurate. We use it as the ground-truth for optimality here.

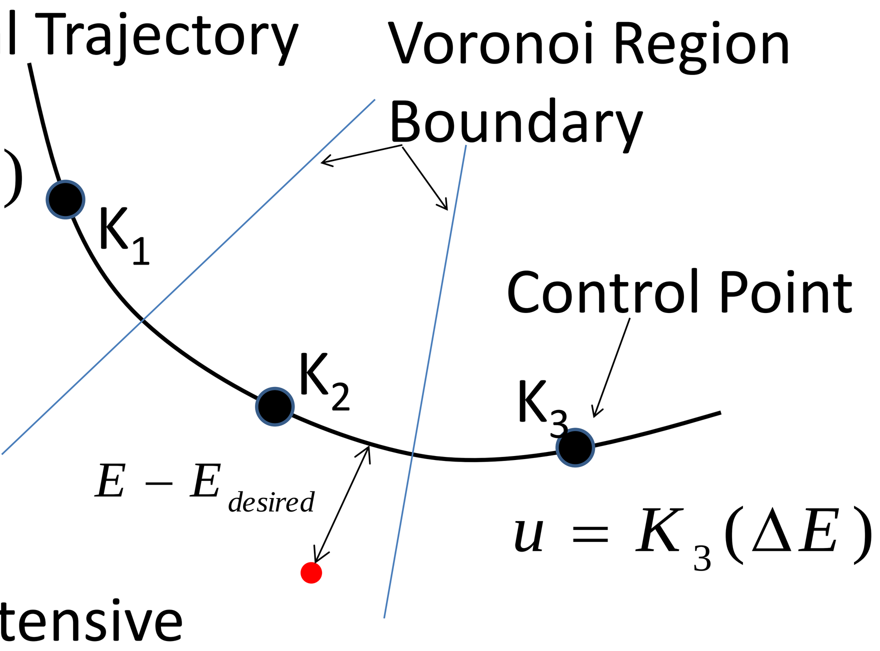
Parametric Control Design Method

1) Create a model
2) Pick a nominal trajectory and define an "Energy" with a constant value $E_{desired}$ on the nominal trajectory
3) Evenly space control points along nominal trajectory
4) Optimize the gains, K_i , for each control point
    - Find gains that result in the lowest possible trajectory cost

Controller:

$$u = K_i(E - E_{desired})$$

Where K_i is the gain associated with the nearest control point.



Conclusion

Advantages:

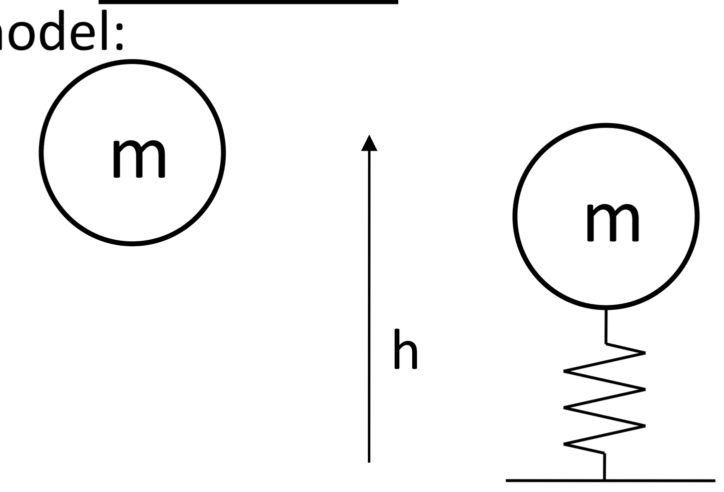
- Simple
- Not computationally intensive
- Performs well near nominal trajectory
- Can be done directly on a real system if necessary (no model)

Disadvantages:

- Requires a nominal trajectory with an invariant quantity
- Difficult to use in higher dimensions

Hopper

1) model:



Flight

$h \geq 0$

$\ddot{h} = -g$

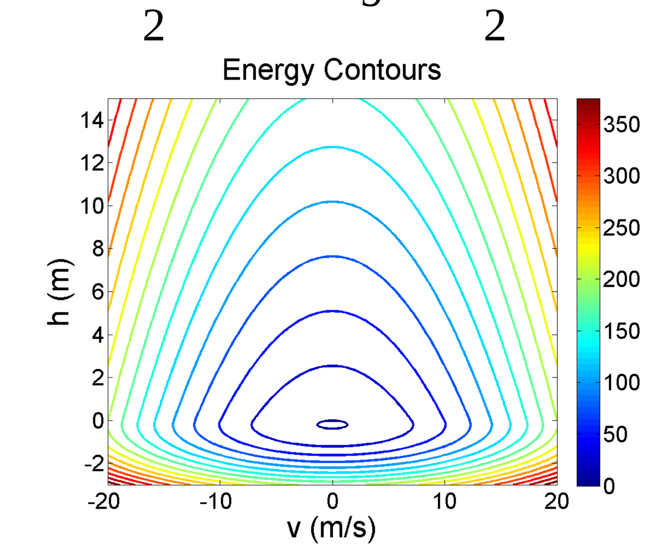
Stance

$h < 0$

$\ddot{h} = \frac{-mg - kh + u}{m}$

2) Kinetic + Potential Energy:

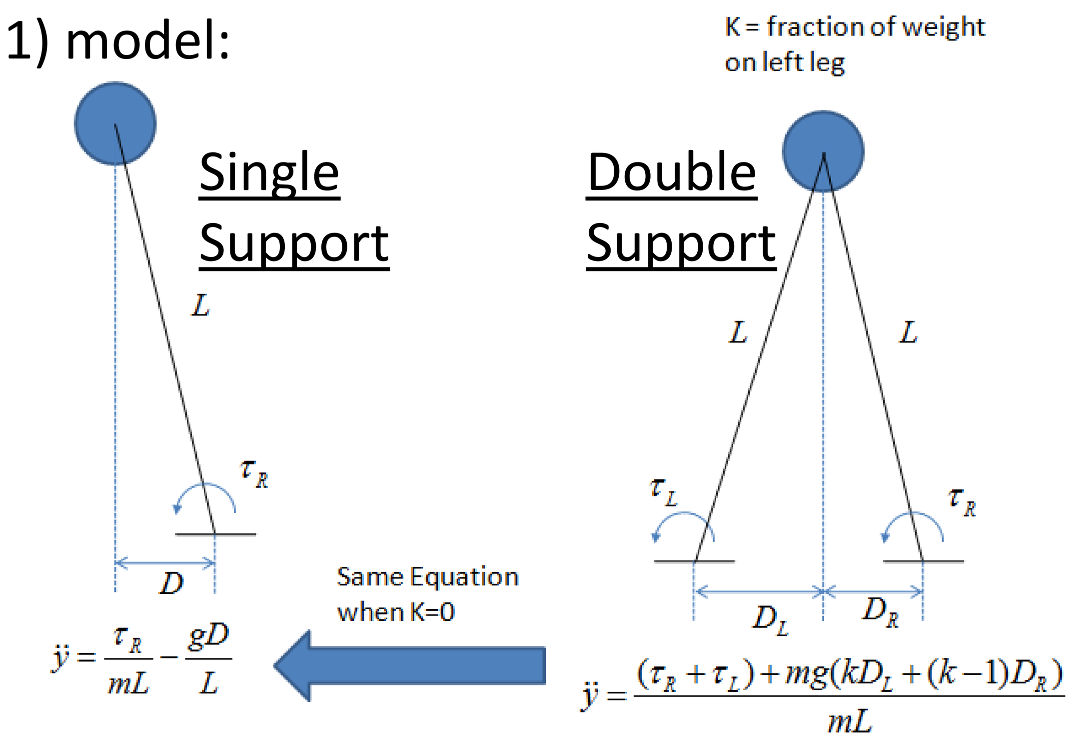
$$E = \frac{1}{2} m\dot{h}^2 + mgh + \frac{1}{2} kh^2$$



3) 15 control points used – spaced evenly along the stance portion of the nominal trajectory

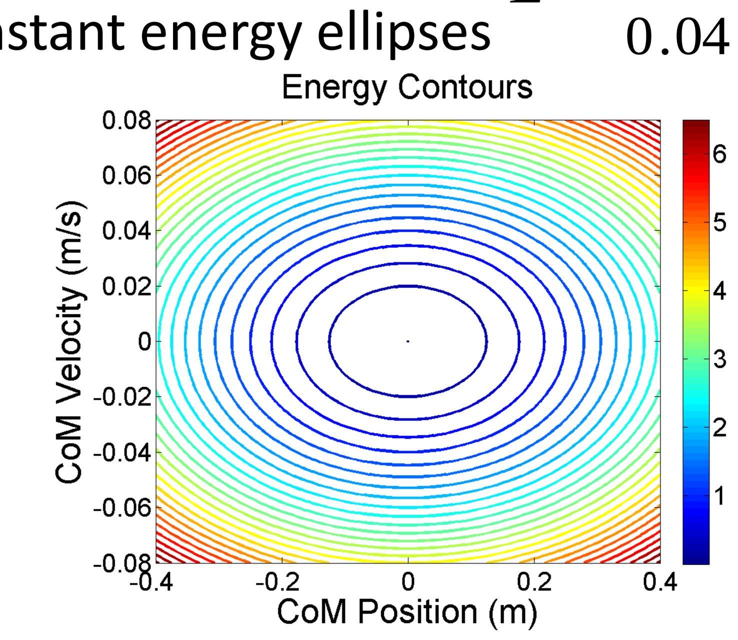
Coronal Plane Biped

1) model:



2) "Orbital Energy": Constant energy ellipses

$$E = \frac{y^2}{0.04^2} + \frac{\dot{y}^2}{0.25^2}$$

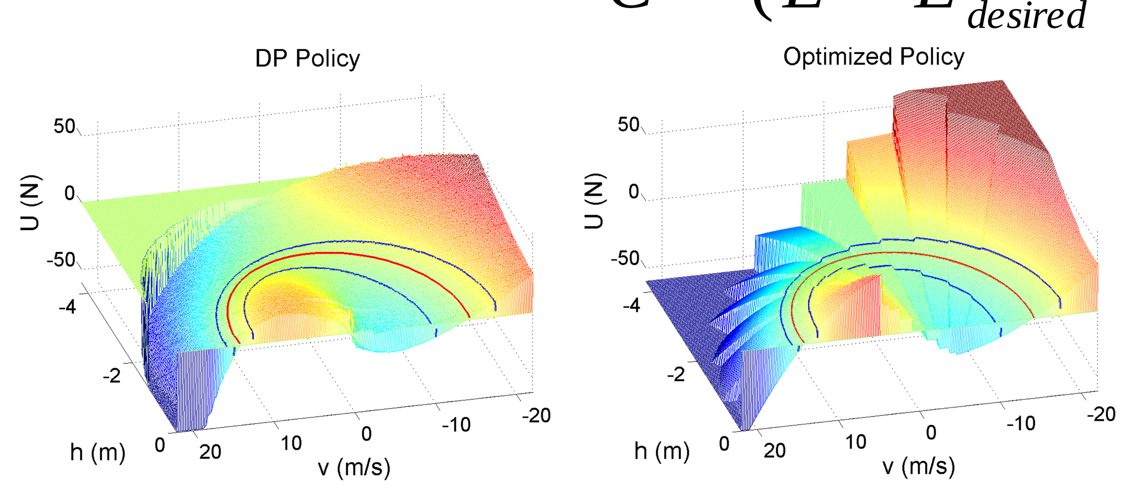


3) 100 control points used

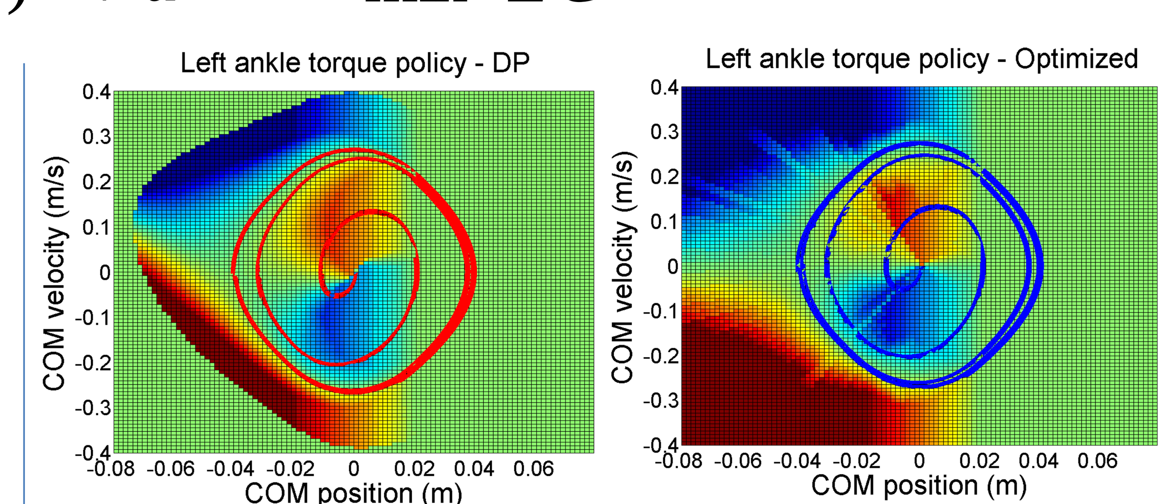
Each wedge of the optimized policy corresponds to 1 control point

4) Optimize gains to minimize error and effort:

$$C = (E - E_{desired})^2 + u^2 \qquad \min \Sigma C$$



Plots of the DP and Parametric policies. Only the stance portion is shown. Nominal trajectory in red. Trajectories starting from ½ and 3/2 desired energies are shown in blue.



Plots of the DP and Optimized policies. Blue means negative torque (push left) and red means positive torque (push right). A trajectory starting from rest is shown.

Trajectory Cost: 960   Trajectory Cost: 1033

7.6% Difference

Result (Hopper):

- Performs well near the nominal trajectory
- Performance degrades drastically far from the nominal trajectory

Result (Coronal Plane Biped):

- Policy and trajectory look qualitatively like those from DP
- Performs well (nearly optimal) even outside of the training range.

  (Trained between ½ and 3/2 of the desired energy; Trajectory started from 0 energy)

This work was funded by the National Science Foundation under grants DGE-0333420, ECS-0325383, and ECS-0824077.