Reinforcement Learning for Dynamically Stable Legged Locomotion

Russell Tedrake & H. Sebastian Seung

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The limit cycle stability of walking robots with only two legs, and even quickly moving robots with four or more legs, cannot be described as transitions between a set of stable fixed points in the configuration space of the robot. In these cases, the stability of the limit cycle depends critically on the dynamics of the walking machine. Because of the complexity of these dynamics, the fact that many of these robots are underactuated[3], and the uncertainty involved with walking on an unmodelled surface, the problem of constructing a control algorithm to provide this limit cycle behavior remains an open and interesting research question.

Approach: While many robotics researchers are working on stability metrics that explicitly incorporate the dynamics of the robot[2], our approach is to formulate dynamic stability as a static stability problem with delayed reward. Using the language of reinforcement learning, the corresponding optimal value function represents a metric of dynamic stability which implicitly incorporates the dynamics of the robot, and the optimal policy yeilds the control algorithm.

Unfortunately, even the simplest example of a dynamically stable legged system[4] presents a daunting task for standard reinforcement learning algorithms, due to the high-dimensional, continuous state and action spaces and the combination of discrete and continuous dynamics. Furthermore, implementations on real robots are limited to a very small number of trials.

Our focus has been on direct policy search using the Pontryagin Minimum Principle[1] evaluated over a large number of initial conditions. Initial experiments have been carried out on a simulation of a simple dynamically stable legged system, yielding promising but computationally expensive results[5]. In particular, we demonstrated a controller for the planar one-legged hopping robot that could recover from a much larger range of initial conditions than the original control algorithms presented in [4]. The important point is that we successfully improved the dynamic stability of the robot by optimizing a cost function which evaluated static stability but was integrated over a finite-horizon. Figure 1 shows a simulation of the robot recovering from an awkward initial condition by spinning its torso 360 degrees around the hip to return to vertical.

Figure 1: The planar one-legged hopping robot is shaded here in progressively darker colors to indicate the passage of time. The initial conditions are represented by the silhouette in the top left corner.

Our current focus in on more elegant solutions to the same problem. This includes attempting to approximate the value function and improving our direct policy search methods using analytical gradient calculations. We are also in the midst of building an experimental platform so that we can demonstrate these ideas on a real robot.
Research Support: This research is supported by the David and Lucille Packard Foundation (contract 99-1471) and by National Science Foundation (grant CCR-0122419).

References:


