Trajectory Optimization and Supervised Learning for Stabilizing Underactuated Bipedal Robot Locomotion

Xingye Da, Ross Hartley, and Jessy W. Grizzle

I. ABSTRACT

For many control tasks, real-time constrained optimization is becoming an important means of designing and implementing feedback control policies. With current computational power, it is not possible to achieve highly dynamic motions (e.g., running or jumping) or to respond to large perturbations with this approach. One alternative is to precompute a set of controllers and build an explicit control policy, such as explicit MPC. However, a sampling method to explore the state space and approximate the control function will suffer the curse of dimension.

This paper proposes an offline approach to design an explicit model-based feedback control policy using ideas from trajectory optimization and Machine Learning (ML). The control design process begins by using trajectory optimization to generate a training set of trajectories that induce periodic at various speeds, both forward and backward, and for various constant ground slopes, flat ground, uphill and downhill. They also include transient gaits that transition among a subset of the periodic gaits in a fixed number of steps. Local controller, such as virtual constraints or foot placement, provides a convenient way to render a local controller for the optimized trajectory.

Supervised learning is then used to train a continuum gait policy from feature to state trajectory. The feature space for the supervised learning includes parameters from a reduced-order biped model (e.g., initial stance leg angle and average speed), exogenous signals (target walking speed is used here, but turning angle could be used as well) and perception input (e.g., terrain height or slope). In experiments, the learned policy allows MARLO to recover from $\approx 200$ N kick. It also enables MARLO to walk down a 22 deg slope and walk on the Wave Field, which presents sinusoidally varying ground height (see Fig. 1).

A. Contributions of the Paper

Our controller has an inner-outer loop structure, such that the inner loop focuses on the actuated variable, hence, it is relatively easier to control. The states used in the outer loop controller, which use the feature in machine learning can be interpreted: in terms of the reduce order model, shown in Fig. 2 Words of caution is also given when update the gait parameter in the outer loop: the stability can be lost when applying supervised learning to train the gait policy even if the inner controller is locally exponentially stable.

Fig. 1: Bipedal robot MARLO walked on the University of Michigan’s Wave Field, a sinusoidally varying grass terrain. Photo was taken by Roger Hart.

Fig. 2: Inner-outer loop control structure. The inner loop is focus on the actuated variable where the outer loop uses low dimensional feature to generate periodic or transient gait.