Optimizing Robotic Single Legged Locomotion with Reinforcement Learning

Péter Fankhauser, Marco Hutter, Christian Gehring, and Roland Siegwart
pfankhauser@ethz.ch
Autonomous Systems Lab, ETH Zurich, Switzerland

1 Motivation

Humans and animals show a remarkable level of proficiency in their ways of locomotion. They exploit the dynamics of the whole body to perform a variety of motions such as jumping and running. The nature of legged robots raises big challenges for controlling these systems. High degrees of freedom (DOF) and highly nonlinear non-smooth dynamics (due to interaction with the environment) count amongst these difficulties. Our goal is to perform highly dynamic jumping and hopping maneuvers with a robotic leg while optimally exploiting the capabilities of the hardware in terms of maximum jump height, jump distance, and energy efficiency. We build on our platform *ScarlETH* [1], an articulated robotic leg which is electrically driven in the hip and knee joints by highly compliant Series Elastic Actuators (SEA) [2].

2 State of the Art

We have seen many different approaches to optimal control of legged robots, e.g. neural networks [3], however, most of which are restricted to simulations. In a closely related work [4], a genetic algorithm is used to evolve a guided vertical jump for a simulated leg with a compliant knee joint. Directly applying the simulation-based trajectories to the physical system is usually unsuitable as models are notoriously difficult to obtain. As a result, a feedback controller is necessary which leads to suboptimal performance as an artificial pattern is forced on the system. As an alternative, reinforcement learning can be applied online on a real robot and promising results have been presented, e.g. in [5]. While classic reinforcement learning algorithms do not scale suitably to high dimensions, recent developments have overcome this limitation with direct learning of a control policy from trajectory rollouts [6].

3 Own Approach

We generate the control policies with reinforcement learning based on the direct policy learning method *Policy Improvement with Path Integrals* (PI^2) [7]. This algorithm has shown to perform well in high-dimensional continuous state spaces and does not rely on the computation of gradients which are sensitive to noise. The control policy is parameterized with gaussian basis functions which are updated in the learning procedure using random exploration rollouts.

We extend the application range of PI^2 from typically slow reaching, grasping and manipulation tasks [8, 9] to highly dynamic maneuvers. We directly parameterize the motor velocity trajectories of the SEA and do not enforce joint position or torque tracking with an additional controller. This way, the learning algorithm learns to excite the inherent dynamics of the system. In order to overcome the model discrepancies, we deploy a combination of simulation and hardware based learning. The simulation allows to quickly converge to a trajectory suitable for the dominating dynamics of the system while the optimization on the real robot compensates for the model inaccuracies.

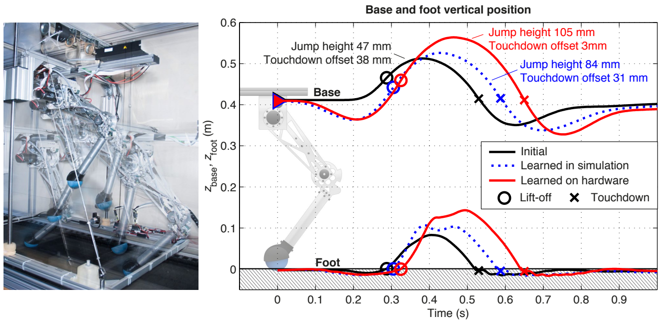

Figure 1: Learning progress of the vertical jump

4 Current Results

We have implemented our approach for different learning tasks. In a first task, the goal is to learn single jumps from a resting posture with maximal height while keeping the touchdown offset (jump distance) to a minimum. The control policy consists of the parameterized desired motor velocity trajectories for hip and knee motor with a total of open 10 parameters. Figure 1 shows the progress of the two step learning framework starting from a manually tuned initial policy (black). While the simulation based policy determined the main form of the trajectory (blue), the optimization on the hardware shows a further significant improvement (red). The algorithm has converged to a characteristic countermovement jump in which the actuators are pre-activated and the body is first lowered to temporarily store energy in the joint springs. This energy is then released during an explosive upwards motion before lift-off.

We have extended the purely vertical jump to learn jumps with a defined height and different distances. By punishing slip of the foot during the thrust phase in the cost function, ScarlETH learns to increase the vertical force on the foot in order to achieve a higher horizontal forces that propel the system to maximal jump lengths. We have created a motion library with different jump lengths and interpolation allows us to reach intermediate lengths with high precision.

In another task, the control parameters of a robust periodic hopping controller are optimized to maximize energy efficiency. Here, we use PI^2 to learn a time-independent control policy with which the algorithm can find an optimal hopping frequency. The algorithm is shaping a non-linear virtual spring characteristic while maintaining the robustness of the controller.

5 Best Possible Outcome

In the future, we will use the presented framework to learn jumping, hopping and running maneuvers on our quadruped robot Starl*ETH* [10].

References

[1] M. Hutter, C. D. Remy, M. A. Hoepflinger, and R. Siegwart, "ScarlETH: Design and control of a planar running robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2011, pp. 562–567.

[2] G. A. Pratt and M. M. Williamson, "Series elastic actuators," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1995, pp. 3137–3181.

[3] J. Helferty and M. Kam, "Adaptive control of a legged robot using an artificial neural network," in *IEEE International Conference on Systems Engineering*. IEEE, 1989, pp. 165–168.

[4] S. Curran and D. E. Orin, "Evolution of a jump in an articulated leg with series-elastic actuation," *2008 IEEE International Conference on Robotics and Automation*, pp. 352–358, May 2008.

[5] R. Niiyama, K. Kakitani, and Y. Kuniyoshi, "Learning to jump with a musculoskeletal robot using a sparse coding of activation," *ICRA*, vol. 0, no. 1, pp. 30–31, 2009.

[6] S. Schaal and C. G. Atkeson, "Learning control in robotics," *IEEE Robotics and Automation Magazine*, no. June, 2010.

[7] E. A. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *The Journal of Machine Learning Research*, vol. 11, pp. 3137–3181, 2010.

[8] J. Buchli, F. Stulp, E. A. Theodorou, and S. Schaal, "Learning variable impedance control," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 820–833, Apr. 2011.

[9] P. Pastor, M. Kalakrishnan, S. Chitta, E. Theodorou, and S. Schaal, "Skill learning and task outcome prediction for manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 3828–3834.

[10] M. Hutter, C. Gehring, M. Bloesch, M. A. Hoepflinger, C. D. Remy, and R. Siegwart, "StarlETH: A compliant quadrupedal robot for fast, efficient, and versatile locomotion," in *Proceedings of the International Conference on Climbing and Walking Robots (CLAWAR)*, 2012.