Feedback Controller for 3D Dynamic Walking using Reinforcement Learning and Hybrid Zero Dynamics

Guillermo A. Castillo [1] **Bowen Weng** [1] **Wei Zhang** [2] **Ayonga Hereid** [3]

1. Abstract

In this work, we present a novel model-free Reinforcement Learning (RL) framework to design feedback controllers for 3D bipedal locomotion. By appropriately incorporating insights from the reduced dimensional representation of Hybrid Zero Dynamics (HZD)-based feedback controllers for 3D robots, we propose a RL framework that yields significantly improved data efficiency, lightweight network structure and short training time. In addition, different from other RL approaches, this method does not depend on prior knowledge of reference trajectories. We demonstrate the effectiveness of the proposed solution to generate a stable walking control policy able to track various walking speeds in different directions on a challenging bipedal robot, Cassie. The controller is robust against external adversarial forces applied at the torso in various directions. Furthermore, this framework presents excellent versatility and generalization due to its independence of a particular robot model.

2. Motivation

3D bipedal walking is a challenging problem due to the multi-phase and hybrid nature of legged locomotion. Properties like underactuation, nonlinear dynamics, ground contacts, and high degrees of freedom significantly increase the model complexity. While model-based methods present formal ways to design feedback control regulators for 3D bipedal walking, these methods are limited by the accuracy of mathematical models to capture the complex dynamics of a 3D robot, which results in non-robust controllers that require additional heuristic compensations and tuning processes. (Grizzle et al., 2014)

Recent progress on deep learning has contributed to the popularity of Reinforcement Learning (RL) for solving challenging control problems in robotics. Existing RL methods often rely on end-to-end learning algorithms that train a neural network (NN) using policy gradient methods, mapping directly the state space to a set of continuous actions (Lillicrap et al., 2015; Schulman et al., 2017; Kidziński et al., 2018). Despite their success, such learning methods are usually sampling inefficient and over-parameterized. Moreover, they may lead to motions unfeasible for real robots and non-smooth control signals.

Through incorporating the insights of Hybrid Zero Dynamics (HZD) with RL training, (Castillo et al., 2019) generated feasible trajectories that are tracked by PD controllers to produce effective walking gaits at different speeds. However, this method only works for a simple 2D robot model. In (Xie et al., 2018), the authors adopts RL methods as part of the feedback control of a 3D bipedal robot. However, the method requires prior knowledge of a good reference trajectory that is used as a based on top of which learned compensations are added to achieve stable walking. An imitation learning approach is applied in (Xie et al., 2019) to a 3D robot, but it also requires a known walking policy that is gradually improved through the learning method.

In this work, we present a novel model-free reinforcement learning framework to design feedback controllers for 3D bipedal locomotion. By appropriately incorporating insights from the reduced dimensional representation of Hybrid Zero Dynamics (HZD)-based feedback controllers for 3D robots, we propose a RL framework that yields a significantly improved data efficiency, lightweight network structure, and short training time. In addition, different from other RL approaches, this method does not depend on prior knowledge of reference trajectories. We demonstrate the effectiveness of the proposed solution to generate a stable walking control policy able to track various walking speeds in different directions on a challenging bipedal robot, Cassie.

3. Problem Formulation

To find feasible trajectories that render stable limit walking cycles, the HZD framework needs to solve an offline optimization problem using the robot's full-order model jointly

[1]Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA. [2]Institute of Robotics, Southern University of Science and Technology, China. [3]Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH, USA. Correspondence to: Guillermo A. Castillo <castillo-martinez.2@osu.edu.ec>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

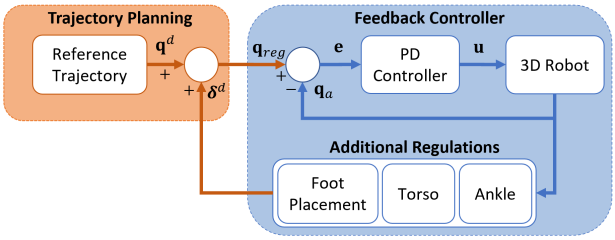

Figure 1. General structure of a traditional HZD-based controller for 3D bipedal walking with heuristic feedback regulations.

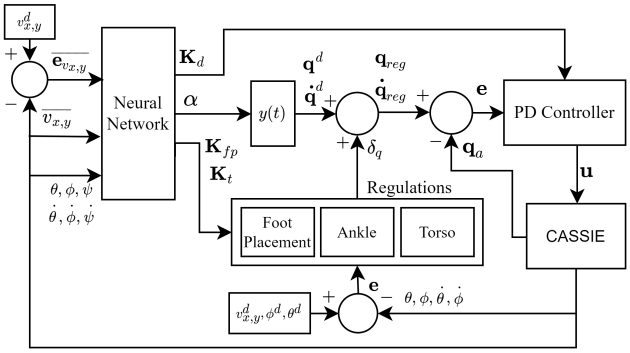

Figure 2. Overall structure of the proposed RL framework, which combines trajectory planning with feedback-based trajectory compensation using additional heuristic regulators.

with virtual constraints that are introduced to synthesize feedback controllers. By designing virtual constraints that are invariant through impact, an invariant sub-manifold –the *hybrid zero dynamics surface*– is created, wherein the evolution of the system is dictated by the reduced-dimensional dynamics of the under-actuated degrees of freedom of the system (Westervelt et al., 2007; Ames, 2014).

However, additional heuristic compensation regulators are often required to compensate for the miss-match between the mathematical model and the real 3D walking robot (Gong et al., 2019; Reher et al., 2016). Usual regulators used in the HZD framework include foot placement, torso regulation, and ankle regulation. Figure 1 shows the scheme of a traditional HZD-based controller with one main block determined by a trajectory planning phase and a second block determined by a feedback controller where the regulations mentioned above are integrated as part of the feedback regulation.

In general, based on feedback information from the robot's torso orientation and velocity, the compensations δ_q will modify the original reference trajectories $\mathbf{q}^d$ to obtain regulated trajectories $\mathbf{q}_{reg}$ that improve the stability and robustness of the walking gaits. Then, the error between the new reference trajectory $\mathbf{q}_{reg}$ and the actual joint trajectories $\mathbf{q}_a$ is used by the low-level PD controllers to produce the control action $\mathbf{u}$, which is translated into the torque applied to the robot joints.

4. Contribution

We propose a non-conventional RL framework that incorporates the physical insight of bipedal walking (symmetric motion, hybrid nature, heuristic compensations) into the control structure and learning process.

A diagram of the overall RL framework is presented in Figure 2, where the trajectory planning stage of Figure 1 has been replaced by a neural network that maps from a reduced order of the robot's state to (i) a set of coefficients α of the Bézier polynomials that define the desired trajectory of the actuated joints, and (ii) a set of gains corresponding to the derivative gain of the joints PD controller $\mathbf{K}_d$, as well as the gains for the foot placement and torso regulations $\mathbf{K}_{fp}, \mathbf{K}_t$.

The neural network implemented for the learning process has 4 hidden layers, each with 32 neurons. The input layer takes the robot's torso desired longitudinal and lateral velocity (v_x^d, v_y^d), average longitudinal and lateral velocity ($\overline{v_x}$ $\overline{v_y}$), average longitudinal and lateral velocity error ($\overline{e_{v_x}}$, $\overline{e_{v_y}}$), roll, pitch and yaw angles (θ, ϕ, ψ), and roll, pitch and yaw angular velocities ($\dot{\theta}, \dot{\phi}, \dot{\psi}$).

The connection between hidden layers is done through ReLU activation functions, and the final layer uses a sigmoid function to limit the range of the outputs. Independent low level PD controllers are then used to track the desired output for each joint, which enforces the compliance of the HZD virtual constraints.

4.1. Learning Procedure

The proposed method can be trained with any RL algorithm that can handle continuous action space, including evolution strategies (ES) (Salimans et al., 2017), deterministic policy gradient methods (Silver et al., 2014), and proximal policy optimization (Schulman et al., 2017). In particular, we used ES for training the network with the following reward function:

$$r = \mathbf{w}^T \mathbf{r}, \tag{1}$$

with a vector of 8 customized rewards $\mathbf{r}$ and the weights $\mathbf{w}$. Specifically,

$$\mathbf{r} = [r_{v_x}, r_{v_y}, r_h, r_u, r_{COM}, r_{ang}, r_{angvel}, r_{fd}]^T. \tag{2}$$

The selected reward encourages velocity tracking (through r_{v_x}, r_{v_y}), height maintenance (r_h), energy efficiency (r_u) and natural walking gaits ($r_{COM}, r_{ang}, r_{angvel}, r_{fd}$).

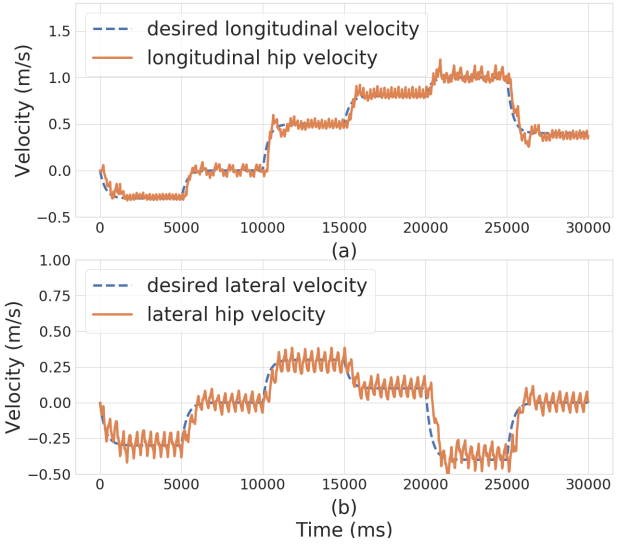

Figure 3. Performance of the learned policy while tracking varying desired longitudinal and lateral walking speeds.

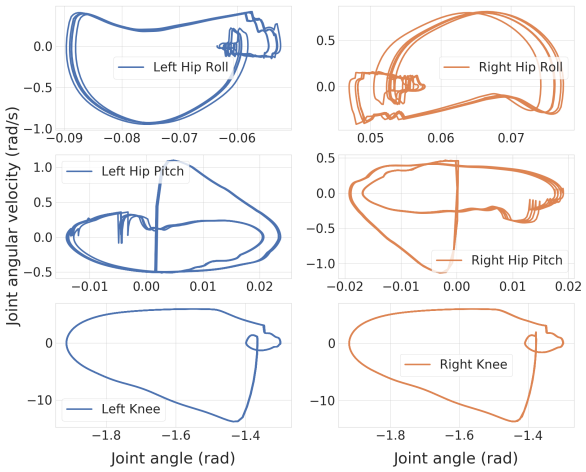

Figure 4. Walking limit cycle of the learned policy with the desired longitudinal velocity of $0.5m/s$.

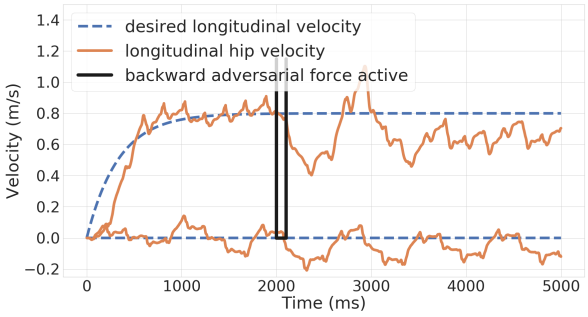

Figure 5. Robustness of the controller when an adversarial force is applied in the forward direction to the robot's pelvis.

5. Simulation Results

To validate the proposed method, a customized environment for Cassie was built using Mujoco (Todorov et al., 2012) and the model provided by Agility Robotics (Agi). The NN has 5069 trainable parameters, and the training time is about 10 hours using a single 12-core CPU machine. The performance of the trained control policy was evaluated in terms of (i) speed tracking, (ii) convergence of stable periodic limit cycles, and (iii) disturbance rejection. The main results are listed below.

5.1. Speed Tracking

Thanks to the decoupled structure of the proposed controller, the learned policy is able to track any desired walking speed and direction within the ranges $[-0.5, 1.0]$ m/s and $[-0.3, 0.3]$ m/s for longitudinal and lateral speed respectively as shown in Figure 3. It is important to denote that the oscillations shown during the tracking of the desired speed are caused by the natural motion described by the torso of the robot during dynamic walking, and that similar effects can be seen in human walking motion and different controllers for bipedal walking robots (Hereid et al., 2014), (Da et al., 2016).

5.2. Stability of the Walking Gait

Figure 4 shows the results when analyzing the stability of the walking gaits realized by the proposed controller. The joints trajectories generated by the trained control policy converge to periodic limit cycles, and the orbit described by corresponding joints in the left and right side are symmetric, showing that the controller realizes stable walking gaits.

5.3. Robustness

The robustness of the controller was tested by applying adversarial forces at the robot's pelvis in the forward and backward directions. The policy handles effectively forces up to 40 N in the forward direction and 45 N in the backward direction without falling or affecting the stability of the walking gait. This result is shown in Figure 5, where a force of 25 N is applied in the backward direction when the robot is walking with a longitudinal velocity of 0.8 m/s and a lateral velocity of 0 m/s.

Finally, we denote that we do not present a comparison of our method with a baseline RL algorithm. The main reason for this is that traditional baseline RL algorithms fail to obtain feasible motions that can be applied on actual bipedal robots. Thus, additional strategies such as imitation learning, iterative learning (Xie et al., 2019), or curriculum-driven learning (Xie et al., 2020) are required to obtain such feasible policies.

References

A simulation library for agility robotics' cassie robot using mujoco. https://github.com/osudrl/cassie-mujoco-sim. Accessed: 2019-09-15.

Ames, A. D. Human-inspired control of bipedal walking robots. *IEEE Transactions on Automatic Control*, 59(5): 1115–1130, May 2014. ISSN 0018-9286. doi: 10.1109/TAC.2014.2299342.

Castillo, G. A., Weng, B., Hereid, A., Wang, Z., and Zhang, W. Reinforcement learning meets hybrid zero dynamics: A case study for rabbit. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 284–290, May 2019. doi: 10.1109/ICRA.2019.8793627.

Da, X., Harib, O., Hartley, R., Griffin, B., and Grizzle, J. W. From 2d design of underactuated bipedal gaits to 3d implementation: Walking with speed tracking. *IEEE Access*, 4:3469–3478, 2016. ISSN 2169-3536. doi: 10.1109/ACCESS.2016.2582731.

Gong, Y., Hartley, R., Da, X., Hereid, A., Harib, O., Huang, J.-K., and Grizzle, J. Feedback control of a Cassie bipedal robot: walking, standing, and riding a segway. *American Control Conference (ACC)*, 2019.

Grizzle, J. W., Chevallereau, C., Sinnet, R. W., and Ames, A. D. Models, feedback control, and open problems of 3D bipedal robotic walking. *Automatica*, 50(8):1955–1988, 2014. ISSN 0005-1098. doi: 10.1016/j.automatica.2014.04.021. URL http://www.sciencedirect.com/science/article/pii/S0005109814001654.

Hereid, A., Kolathaya, S., Jones, M. S., Van Why, J., Hurst, J. W., and Ames, A. D. Dynamic multi-domain bipedal walking with atrias through slip based human-inspired control. In *Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control*, HSCC '14, pp. 263–272, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2732-9. doi: 10.1145/2562059.2562143.

Kidziński, Ł., Mohanty, S. P., Ong, C. F., Huang, Z., Zhou, S., Pechenko, A., Stelmaszczyk, A., Jarosik, P., Pavlov, M., Kolesnikov, S., Plis, S., Chen, Z., Zhang, Z., Chen, J., Shi, J., Zheng, Z., Yuan, C., Lin, Z., Michalewski, H., Milos, P., Osinski, B., Melnik, A., Schilling, M., Ritter, H., Carroll, S. F., Hicks, J., Levine, S., Salathé, M., and Delp, S. Learning to run challenge solutions: Adapting reinforcement learning methods for neuromusculoskeletal environments. In Escalera, S. and Weimer, M. (eds.), *The NIPS '17 Competition: Building Intelligent Systems*, pp. 121–153, Cham, 2018. Springer International Publishing. ISBN 978-3-319-94042-7.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *CoRR*, abs/1509.02971, 2015.

Reher, J. P., Hereid, A., Kolathaya, S., Hubicki, C. M., and Ames, A. D. Algorithmic foundations of realizing multi-contact locomotion on the humanoid robot DURUS. In *the 12th International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, San Francisco, December 2016. Springer. URL http://wafr2016.berkeley.edu/.

Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. Deterministic policy gradient algorithms. 2014.

Todorov, E., Erez, T., and Tassa, Y. MuJoCo: a physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, October 2012. doi: 10.1109/IROS.2012.6386109.

Westervelt, E. R., Grizzle, J. W., Chevallereau, C., Choi, J. H., and Morris, B. *Feedback control of dynamic bipedal robot locomotion*. CRC press Boca Raton, 2007.

Xie, Z., Berseth, G., Clary, P., Hurst, J., and van de Panne, M. Feedback control for cassie with deep reinforcement learning. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1241–1246. IEEE, 2018.

Xie, Z., Clary, P., Dao, J., Morais, P., Hurst, J., and van de Panne, M. Iterative Reinforcement Learning Based Design of Dynamic Locomotion Skills for Cassie. *arXiv e-prints*, art. arXiv:1903.09537, Mar 2019.

Xie, Z., Ling, H. Y., Kim, N. H., and van de Panne, M. Allsteps: Curriculum-driven learning of stepping stone skills. 2020.