Learning an Efficient Walking Trot using Variable Impedance

Elco Heijmink[1,2] Andreea Radulescu[2], and Claudio Semini[2]

Abstract— **The increased complexity of the tasks that robots can perform nowadays, increases the required tuning of gait defining parameters as well. In this work we propose a methodology to learn a variable impedance strategy with minimal need of expert knowledge and a short learning time. We show results validating the strategy on a set of terrains and for various speeds using impedance control on a hydraulic quadruped.**

I. INTRODUCTION

Modern robotic platforms can achieve highly complicated tasks, such as running, balancing and locomotion over challenging terrain. Despite this, in real world scenarios, biological systems can still outperform most robots [7]. The successful execution of a complex robotic task, usually requires tuning of a large number of parameters like the gains of the controller and gait defining values. This typically is a lengthy empirical procedure, performed by an expert user.

Additionally, modulating compliance during locomotion can improve the behaviour on factors like the energy efficiency [9]. This is inspired by the performance of animals, which have been shown to use a impedance strategy in their behaviour [3]. Defining an optimal impedance profile for a robotic legged system is not trivial. Reinforcement learning provides us with a feasible strategy determining such a profile alongside other parameters of interest, thus avoiding the requirement to hand-tune these values.

Previous studies showed that The Policy Improvements with Path Integrals (PI^2) reinforcement learning algorithm [8] is an effective way to simultaneously optimise gait parameters and an impedance profile [2]. The PI^2 requires no tuning of algorithmic parameters besides the exploration noise. This makes it a good candidate for creating a practical methodology to optimise the locomotion without the need of expert knowledge.

Learning algorithms can require a large number of trials over a long time in order to find the optimal desired behaviour. Hence, the learning time becomes a critical factor in the ability to tackle large numbers of scenarios.

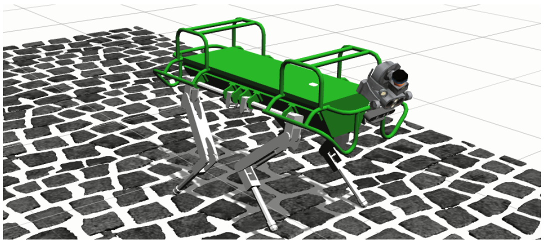

Fig. 1. HyQ trotting on cobblestones during learning in simulation.

In this work, we improve on the methodology introduced by Ponton et al. [5], which optimises a trotting gait by modulating the end effectors' impedance. We augment the abilities of the procedure by significantly improving the learning time (from over 90 minutes to under 10 minutes). The procedure is now designed to run as a parallel process on the background, while the robot performs a main task. We also focus on a previously unexplored application of the approach, by setting the main goal on energy efficiency maximisation. We present results obtained on a realistic stimulation of the hydraulic torque controlled quadruped HyQ [6].

II. TECHNICAL APPROACH

We aim to optimise the energy efficiency of the quadruped by learning a variable impedance strategy for a walking trot, while maintaining stability. To this aim three sets of *learning variables* are optimised: i) gait parameters (i.e., step height, step length, stride frequency, duty factor), ii) the gains for the control architecture, and iii) the end effectors' impedance profiles. The control architecture will be elaborated in Section III.

The PI^2 algorithm used learns based on the iterative rollouts of the given learning variables. These variables are generated by injecting zero-mean Gaussian noise into the current learning policy. We collect data on the trotting behaviour generated by these parameters (approx. 5 to 15 seconds), which is evaluated using the cost function. The current learning policy is updated using a sum of the eight best learning variables, weighted by their costs.

The cost function consists of three main terms: the speed tracking error, the energy consumption and the motion of the trunk. Additionally, costs enforcing the joint limits and penalising high torques are added to avoid infeasible solutions. The speed tracking is expressed as the sum of the instantaneous speed tracking errors. The energy consumption is expressed in the form of the Cost of Transport (CoT). The metric for the trunk motion is elaborated in the next paragraph.

As previously mentioned one of the challenges of the learning routine is the optimisation time. To reduce the evaluation time of the rollouts, the cost function is constructed based on stable values. Instead of using the noisy direct measurement of the trunk motion (i.e. the roll and pitch angles), a stable estimator is used. This trunk motion

---

[1] BioMechanical Engineering, Faculty of Mechanical Engineering, Delft University of Technology, Mekelweg 2, Delft 2628 CD, The Netherlands `E.Heijmink@student.tudelft.nl`

[2] Department of Advanced Robotics, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy. `firstname.lastname@iit.it`

estimator is chosen as the distance between the Centre of Pressure (CoP) and the Centre of Mass (CoM). This value shows a strong correlation with the peaks of the roll and pitch of the trunk, but is more steady. Using this estimator in the cost function for penalising extensive trunk motions improves the disturbance reaction calculated using the Gait Sensitivity Norm (GSN). This is a disturbance rejection measure proposed in [4], which was proven to show a good correlation with the real disturbance rejection in limit cycle walkers and can be calculated from the observed behaviour.

To reduce the number of rollouts needed for the cost function convergence, the number of learning variables has been kept to a minimum. This was mainly achieved by learning the impedance only for the swing phase including the touchdown and liftoff, while using a constant impedance value for the stance phase. This is due to the main advantages of the optimisation coming from the effect of impedance modulation during swing time.

III. SIMULATION EXPERIMENTS AND RESULTS

To evaluate the method proposed in this work, we use the rigid structured quadruped HyQ [6] (Fig. 1). The control architecture is based on the Reactive Controller Framework (RCF)[1], which generates elliptical trajectories for the feet of the robot. An additional trunk controller maintains the attitude of the robot, compensating deviations in the roll and pitch angles, while mapping the stabilisation forces into joint torques without influencing the joint trajectories. The gains of this trunk controller are optimised by learning, as mentioned in Section II.

The proposed learning algorithm was tested in three environments: i) flat terrain, ii) rough terrain with cobble stones (as depicted in Fig. 1) and iii) a terrain consisting of 5° ramps with a maximum height variation of $17.5cm$. The desired forward velocity is varied between $0.3m/s$ and $0.8m/s$.

The resultant behaviour shows an improvement in terms of energy efficiency varying from 10% on flat to 21% on rough terrain, compared to the performance of a hand-tuned strategy, as currently employed on the platform. We validate the stability of the obtained strategy by employing the GSN as described in Section II. On this metric, we observed an improvement in the disturbance rejection. An example of the evolution of the cost function during one of these experiments (i.e., the cobblestone terrain) is shown in Fig 2. The computation time of the procedure varies between 5 and 10 minutes. This is a significant improvement compared to a similar procedure presented in [5] which requires at least 90 minutes.

Much of this improvement has been achieved by learning only the swing phase impedance of the trot limit cycle. Our preliminary analysis indicated that learning the full cycle leads up to a doubling of the computation time. On the other hand 80% of the overall energy efficiency improvement can be achieved just by learning the impedance during the swing phase including the touchdown and liftoff.

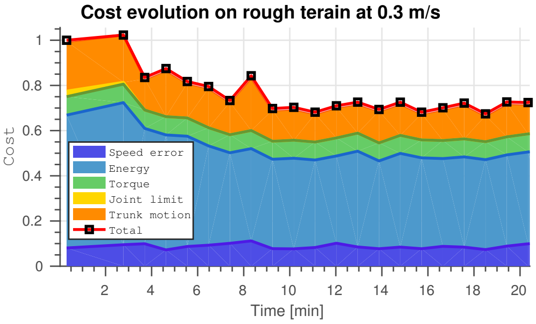

Fig. 2. Evaluation of cost function on the rough terrain with cobble stones, each square represents a parameter update.

IV. CONCLUSIONS

We presented a methodology that can improve the efficiency of a trotting gait by learning an impedance modulation strategy. Compared to passive compliance elements impedance control lacks the ability to store energy. We show that without this ability, impedance control can still increase the energy efficiency. We hypothesize that impedance control exploits other advantages like the increased robustness towards disturbances which will increase the energy efficiency. The method proved to be efficient on a set of various terrains and speeds. By using a dedicated cost function and limiting the number of optimisation parameters, we improve the learning time, compared to previous approaches.

In future work we aim to expand this approach by learning the impedance profile during the whole execution. Since this has been shown to significantly increase the required learning time, new strategies to overcome this difficulties will have to be devised. In future work we aim to address these issues, as well as performing more extensive tests using the available hardware platform.

REFERENCES

[1] V. Barasuol, J. Buchli, C. Semini, M. Frigerio, E. R. De Pieri, and D. G. Caldwell. A reactive controller framework for quadrupedal locomotion on challenging terrain. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2554–2561, 2013.

[2] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal. Learning variable impedance control. *The International Journal of Robotics Research*, 30(7):820–833, 2011.

[3] H. M. Herr and T .A. McMahon. A galloping horse model. *The International Journal of Robotics Research*, 20(1):26–37, 2001.

[4] D. G.E. Hobbelen and M. Wisse. A disturbance rejection measure for limit cycle walkers: The gait sensitivity norm. *IEEE Transactions on robotics*, 23(6):1213–1224, 2007.

[5] B. Pontón, F. Farshidian, and J. Buchli. Learning compliant locomotion on a quadruped robot. In *IROS Workshop (Ed.), Compliant manipulation: Challenges in learning and control*, 2014.

[6] C. Semini, N. G. Tsagarakis, E. Guglielmino, M. Focchi, F. Cannella, and D. G. Caldwell. Design of hyq–a hydraulically and electrically actuated quadruped robot. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 225(6):831–849, 2011.

[7] Vance A T. The energetic cost of moving. *American Scientist*, 63(4):413–419, 1975.

[8] E. Theodorou, J. Buchli, and S. Schaal. Learning policy improvements with path integrals. In *AISTATS*, pages 828–835, 2010.

[9] B. Vanderborght, B. Verrelst, R. Van Ham, M. Van Damme, P. Beyl, and D. Lefeber. Development of a compliance controller to reduce energy consumption for bipedal robots. *Autonomous Robots*, 24(4):419–434, 2008.