Reinforced Learning for Ground Movement of the Hyperreduntant Modular Robot

Jacek Szklarski[0000–0003–1541–9878], **Ela Zawidzka**[0000–0003–1243–9355], **Machi Zawidzki**[0000–0001–8695–4400]

Institute of Fundamental Technological Research
Polish Academy of Sciences
Pawińskiego 5b, 02-106 Warsaw, Poland
jszklar@ippt.pan.pl

Abstract. *We show that the Soft Actor-Critic Reinforced Learning algorithm is able to find efficient motion patterns for a hyper-redundant robot consisting of 6 identical modules connected in a chain like fashion. The control is done by applying relative angular velocities between the modules. Analogous system has been studied before in the context of a robotic trunk-like manipulator.*
Keywords: *Reinforced Learning, Robotic Locomotion, Hyperreduntant Modular Manipulator, Arm-Z, Extremely Modular System*

1. Introduction

Biological snakes are extremely well adapted for different environments. This is mostly the result of the high redundancy of the snake mechanisms. In many instances of irregular environments the bio-inspired robots outperform conventional wheeled, legged or tracked robots. The snake-resembling robots are researched already for a few decades. This type of locomotion has been studied already in the 1940s [1], and a half century later, its rigorous mathematical model has been developed. In the late 90's, a trunk-like locomotors and manipulators have been introduced in [2].

A number of various snake-like robots have been built [3]; most of the designs were intended for crawling on ground [4, 5, 6, 7, 8], some of them for swimming [9, 10], and even fewer for both swimming and crawling on the ground [11, 12].

Snake-like robots are also developed in the context of space applications. In particular, NASAs Exobiology Extant Life Surveyor (EELS) [13] – somehow mechanically similar to the robot discussed here – is being designed for operation in icy crust of Saturn's moon Enceladus (however, EELS, has many potential down-to-Earth application as well).

Fig. 1 shows one of modular snake robots developed in Biorobotics Lab at the Carnegie Mellon University, which is not only capable of crawling on the ground but can also climb vertical posts (for more information see [14]).

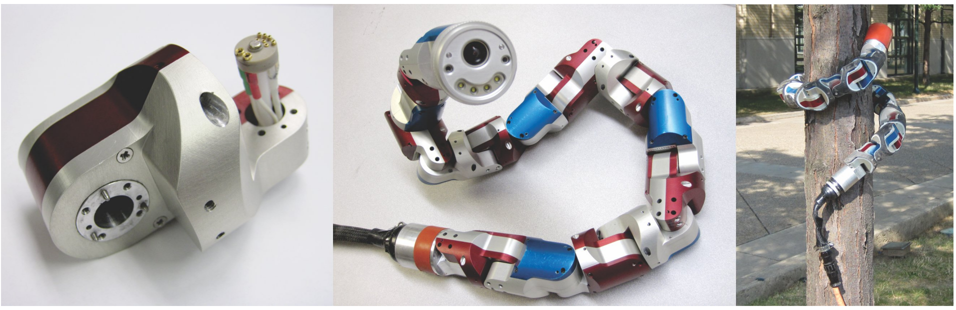

Figure 1. Modular Snake Robot "Uncle Sam". From the left: the module, the robot and the tree climbing action. Photographs ©2010 Biorobotics Lab at the Carnegie Mellon University.

For a concise overview of other biological inspirations for robot concepts see [15].

2. Extremely Modular Hyperredundant Robot

The snake robot presented here is based on *Arm-Z*, which is a concept of robotic manipulator introduced in [16] composed of congruent modules each having one degree of freedom (1-DOF) - a relative twist. In principle, each module is defined by the following parameters: size r, offset d, and ζ, that is the angle between upper (**T**) and lower (**B**) faces of the module. s (slenderness), is an additional parameter, i.e. a d to r ratio. Fig. 2 shows the geometrical interpretation of these parameters and the early, functional prototype.

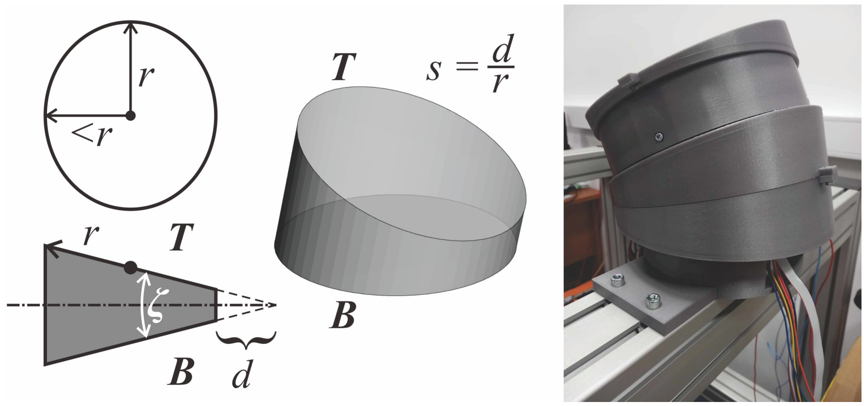

Figure 2. On the left: visualization of the *Arm-Z* unit defined by three parameters: r, d and ζ. On the right: an early functional prototype of the *Arm-Z* manipulator with partial chain of congruent units.

For more information on the concept of *Arm-Z* including its folding study see [17], for early successful implementations of meta-heuristics to control of *Arm-Z* see [18], for documentation of several preliminary prototypes see [19, 20]. The domain of robotic locomotion traditionally uses classical methods designed to address the intricate challenges of enabling robots to navigate their surroundings. Conventional approaches involve specifically engineered control systems and predefined algorithms. However, despite these classical methodologies, locomotion in complex and dynamic environments remains difficult due to the inherent uncertainties and mathematical problems related to kinematic and inverse kinematics. The difficulties in modeling and adapting to diverse terrains, unforeseen obstacles, and dynamic scenarios underscore the need for innovative solutions. Reinforcement learning (RL) emerges as a compelling alternative, offering the capability to learn adaptive locomotion strategies through interaction with the environment and it has been successfully applied in many cases [21]. Due to the non-trival kinematics and possible ways of interaction with the environment, it is tempting to verify if RL can be used to control the hyper-redundant manipulator so as to enable its horizontal movement on the ground.

3. The model and results

Choosing the right physics engine is of fundamental importance for effective RL in robotics. The physics engine acts as the virtual playground for agents to learn and interact, impacting training accuracy and efficiency. For the results presented here, PyBullet [22] was use due to its accuracy in modeling physical interactions, computational efficiency, and open source nature. The mass of each module is assumed to be 1kg and the radius $R = 1m$. The friction parameters for the flat ground are lateral=1.0, rolling=0.5, spinning=0.9. Using these values PyBullet calculates “real physical” coefficients of friction during contact.

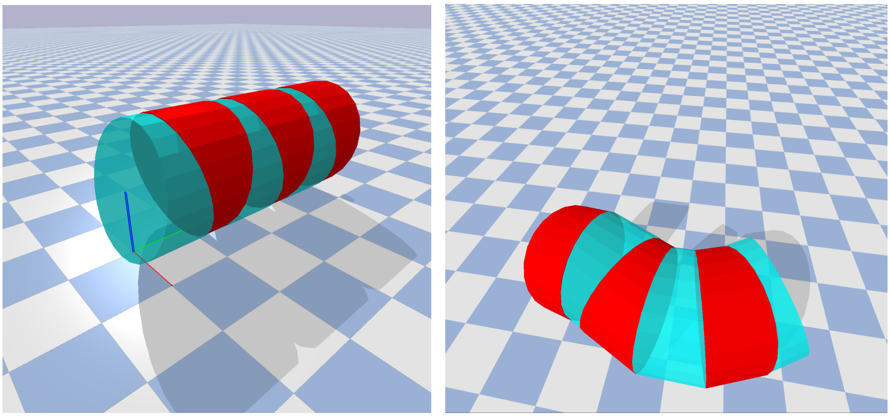

Figure 3. Images rendered by the physics engine representing the robot (alternating colors were chosen to distinguish the modules). Left: the initial state, right the state at timestep=70.

States of the system is described by a vector $S = [s_1, \ldots, s_5]$ in this case, since there are 6 modules and therefore 5 relative twists between them, described as s_i [rad]. The observation of the model is assumed to be simply S. Actions consist of expected angular velocities $A = [\omega_1, \ldots \omega_5]$, the maximum allowed velocity is ± 1rad s^{-1}. The physics engine applies torque in order to achieve the desired angular velocity, the maximum allowed value of torque is 100 [N m].

As the RL algorithm, the Soft Actor-Critic (SAC) off-policy [23] was chosen in the discussed case and its implementation in Stable-Baselines3 [24]. SAC has been successfully applied to a number of interesting problems, including navigation of mobile robots [25]. After basic hyperparameter tuning, the following parameters were applied for learning: learning rate=7.3e-4, batch size=256, γ=0.98, τ=0.02, train freq=8, and SDE were used. The same network architecture (256,256) was used for both the actor and the critic. At each step the physics engine simulates 0.5s during which the torques are applied and contact points and frictions are handled. The reward is simply the sum of the distanced of module centers from the center of the coordinate system. It rewards strategies generating fast motion in a straight line. Initially the robot lies flat $s_i = \pi$ along the y axis (Fig. 3 left). During the learning process the assumed maximal time horizon is 200 time steps. After about 100k steps of training the model is trained in the sense that the average return does not improve. The model described below is the best in 1M steps (5k episodes).

Figure 4 depicts actions and states in the first 100 steps generated by the model from the initial state. It is evident that - after the initial transition - a rather simple velocity pattern emerges which basically constitutes of two configurations. Switching between these two states leads to motion along almost straight line.

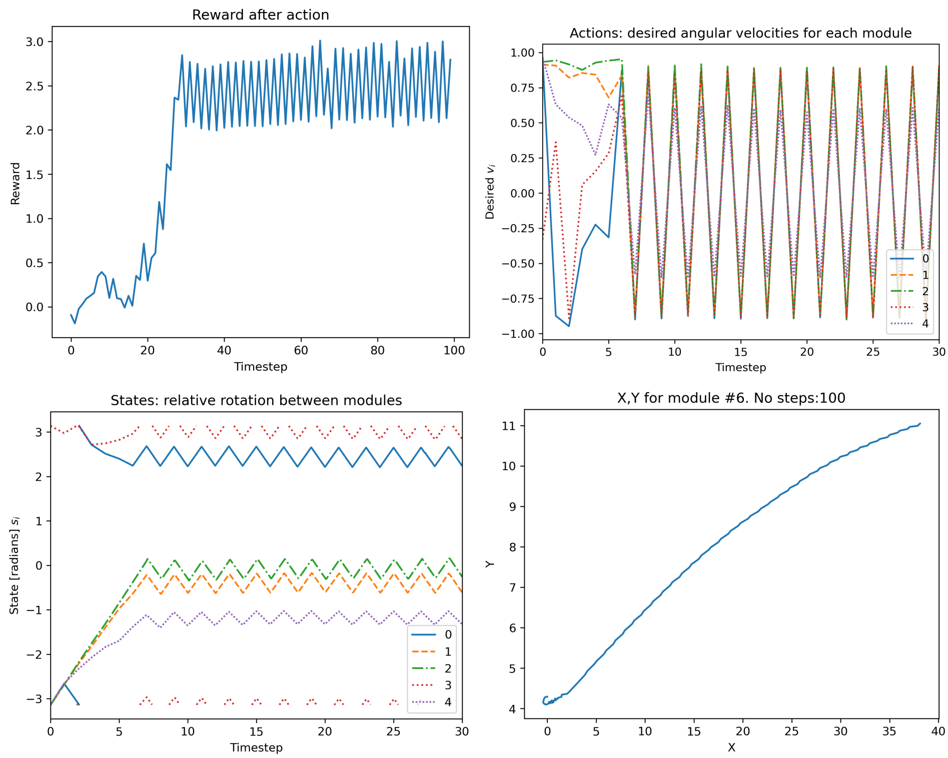

Figure 4. Reward, corresponding actions ω_i, states s_i and a the projection of the module 6 center onto the (x, y) plane during first 30 or 100 timesteps (corresponding to 15 or 50 s).

4. Conclusions

It has been shown that – by means of RL methods – it is possible to obtain motion patterns generating horizontal movement on the ground of the discussed robot. Further study will focus on models with larger number of modules as well as incorporating more control regarding movement direction. Study against module failure will be performed. Current simulations lack any random fluctuations of the environment (except the random nature of the RL control algorithms) which might have significant impact on the effectiveness of the RL algorithms. Our ongoing research is focused on learning with simulations in more complex, random, hilly terrain and additional noisy perturbations in the control process. Additionally, work on physical prototype of the robot will be continued which will also make possible to perform simulations for RL with more realistic physical properties.

Acknowledgments

This research is a part of the project titled *Arm-Z: an extremely modular hyper-redundant low-cost manipulator – development of control methods and efficiency analysis* and funded by *OPUS 17* research grant No. 2019/33/B/ST8/02791 supported by the National Science Centre, Poland.

References

[1] Gray, J. The mechanism of locomotion in snakes. *J Exp Biol.*, 23(2):101–120, 1946.

[2] Hirose, S. *Biologically Inspired Robots: Snake-Like Locomotors and Manipulators*. Oxford University Press, 1993.

[3] Ijspeert, A. J. and Crespi, A. Online trajectory generation in an amphibious snake robot using a lamprey-like central pattern generator model. In *Proceedings of the 2007 IEEE international conference on robotics and automation (ICRA 2007)*, pages 262–268. IEEE, 2007.

[4] Chirikjian, G. and Burdick, J. Design, implementation, and experiments with a thirty-degree-of-freedom "hyper-redundant" robot. *ISRAM 1992*, 1992.

[5] Klaassen, B. and Paap, K. L. GMD-SNAKE2: a snake-like robot driven by wheels and a method for motion control. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, volume 4, pages 3014–3019. IEEE, 1999.

[6] Miller, G. Snake robots for search and rescue, neurotechnology for biomimetic robots. In *Proc. Int. Conf. Robotics and Automation (ICRA)*, pages 16–21. 2002.

[7] Choi, H. and Ryew, S. Robotic system with active steering capability for internal inspection of urban gas pipelines. *Mechatronics*, 12(5):713–736, 2002.

[8] Tsakiris, D. P., Sfakiotakis, M., Menciassi, A., La Spina, G., and Dario, P. Polychaete-like undulatory robotic locomotion. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 3018–3023. IEEE, 2005.

[9] Melsaac, K. and Ostrowski, J. P. A geometric approach to anguilliform locomotion: modelling of an underwater eel robot. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, volume 4, pages 2843–2848. IEEE, 1999.

[10] Wilbur, C., Vorus, W., and Cao, Y. 14 a lamprey-based undulatory vehicle. *Neurotechnology for biomimetic robots*, 1(9):285, 2002.

[11] Yamada, H. Development of amphibious snake-like robot ACM-R5. In *the 36th International Symposium on Robotics (ISR 2005), Tokyo*. 2005.

[12] Crespi, A., Badertscher, A., Guignard, A., and Ijspeert, A. J. AmphiBot I: an amphibious snake-like robot. *Robotics and Autonomous Systems*, 50(4):163–175, 2005.

[13] Ono, M., Carpenter, K., Cable, M. L., Wilcox, B. H., and Tosi, L. P. Exobiology extant life surveyor (eels). In *AGU Fall Meeting Abstracts*, volume 2019, pages P21D–3410. 2019.

[14] Wright, C., Johnson, A., Peck, A., McCord, Z., Naaktgeboren, A., Gianfortoni, P., Gonzalez-Rivero, M., Hatton, R., and Choset, H. Design of a modular snake robot. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2609–2614. IEEE, 2007.

[15] Zielińska, T. Transforming biological patterns into robot concepts. *Pomiary Automatyka Robotyka*, 27, 2023.

[16] Zawidzki, M. and Nagakura, T. Arm-z: a modular virtual manipulative. In *Proceedings of the 16th International Conference on Geometry and Graphics*, pages 75–80. 2014.

[17] Zawidzki, M. Deployable pipe-z. *Acta Astronautica*, 127:20–30, 2016.

[18] Zawidzki, M. and Szklarski, J. Transformations of Arm-Z modular manipulator with Particle Swarm Optimization. *ADV ENG SOFTW.*, 126:147–160, 2018.

[19] Zawidzka, E., Kiński, W., and Zawidzki, M. Preliminary prototype of a 4-unit arm-z hyper-redundant modular manipulator. In *Conference on Automation*, pages 285–294. Springer, 2021.

[20] Zawidzka, E., Szklarski, J., Kiński, W., and Zawidzki, M. Prototype of the arm-z modular solar tracker. In *Conference on Automation*, pages 273–282. Springer, 2022.

[21] Zhang, T. and Mo, H. Reinforcement learning for robot research: A comprehensive review and open issues. *International Journal of Advanced Robotic Systems*, 18(3):17298814211007305, 2021.

[22] Coumans, E. and Bai, Y. Pybullet, a python module for physics simulation for games, robotics and machine learning. `http://pybullet.org`, 2016–2021.

[23] Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018.

[24] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021. URL `http://jmlr.org/papers/v22/20-1364.html`.

[25] de Jesus, J. C., Kich, V. A., Kolling, A. H., Grando, R. B., Cuadros, M. A. d. S. L., and Gamarra, D. F. T. Soft actor-critic for navigation of mobile robots. *Journal of Intelligent & Robotic Systems*, 102(2):31, 2021.