CALL FOR APPLICATIONS FOR A MASTER INTERNSHIP

ZERO-ORDER TRAJECTORY OPTIMIZATION

ABSTRACT

Optimal control is a way to program robots by defining the tasks to be achieved in terms of quantitative objectives such as cost, reward or constraint functions, rather than by explicitly programming the motion by sequences or demonstrations. Among the properties that characterize the methods used to solve it, the most important are often whether the decision variable that is optimized is the robot trajectory or its policy; and whether this optimization uses derivatives or not. While trajectory optimization often benefits from the derivatives, it also makes it less robust in particular when considering irregular problems such as movements with contacts or with integer decision. In this project, it is proposed to consider gradient-free algorithms for optimizing the robot motion, in particular to compare it to gradient-based optimization, to seek for guarantees in the convergence (such as rate of convergence or convergence domain), to transfer some algorithmic progress developed in gradient-based trajectory optimization, and to understand the importance or limitations of not using the gradient in reinforcement learning.

PROJECT

The objective of this project is to understand the importance of derivatives when optimizing the movement of a robot. While there is a strong understanding that predictive control (aka trajectory optimization) and reinforcement learning (aka policy optimization) are two approaches to compute the unique solution of an optimal control problems [1], algorithms to tackle both are very different in practice. Predictive control implies quick optimization at run time, which is mostly achieved using strong models implementing the derivatives of the objective functions and the robot simulation [2]. This enables the solver to quickly (super-linearly converge), but limits the use to mostly regular problems. By solving reinforcement learning off-line on multicore processing units, the importance of the algorithmic efficiency is relaxed (actually, we barely have any guarantees on the convergence rate) which enables the use of gradient free optimization, easier to implement when the system is not smooth [3].

There are some few research actions in the world to understand this duality. Some groups highlight the importance of the randomization in approximating the gradients [4,5], other seeks for exploiting the gradient in reinforcement learning [6]. With this project, we propose to reverse the study by focusing on the properties we could get from gradient-free trajectory optimization [8,9]. Some early tentative have been proposed 10 years ago [10,11], yet limited by the capabilities of CPU at this time. More recently, new algorithms have been proposed combining zero-order evaluations with local reconstruction approximating the local landscape and leading to guaranteed convergence rate. More pragmatic algorithms have been explored taking into account the architecture of available processing units. We propose to combine these ideas while exploiting the structure of the trajectory optimization problem, as done in our recent gradient-based solver. We expect that prototype could significantly advance the state of the art in solving off-line locomotion problems such as stair climbing and manipulation problem such as re-arrangement.

The understanding gained in studying zero-order trajectory optimization, especially for nonsmooth problems, could then be matched with randomized gradient estimators and gradient-free reinforcement learning, in particular by proposing a new version of the classical guided-policy search suitable to any class of problems.

CONTEXT

This internship will take place at LAAS-CNRS in the Gepetto team in Toulouse. It will be integrated to the EU project Agimus, in close collaboration with Inria Paris and CTU Prague, among other partners.

contact nmansard@laas.fr

REQUIREMENTS

- MSc student, preferably from Computer Science or Applied Mathematics
- Mathematics, control theory, robotics or computer vision background is desirable
- The implementation will be done in Python for the prototype, C++ for a possible second more efficient version. Interest in software development in expected.
- Prior courses, knowledge, or practical interest in robotics, machine learning, or numerical optimization is a plus

LAAS CNRS

REFERENCES

[1] Song, Y., Romero, A., Müller, M., Koltun, V., & Scaramuzza, D. (2023). Reaching the limit in autonomous racing: Optimal control versus reinforcement learning. Science Robotics, 8(82), eadg1462. <I was not able to find a free PDF for this one. Maybe I should not cite it!>

[2] Tassa, Y., Erez, T., & Todorov, E. (2012, October). Synthesis and stabilization of complex behaviors through online trajectory optimization. In IEEE/RSJ IROS 2012. https://homes.cs.washington.edu/~todorov/papers/TassaIROS12.pdf

[3] Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V., and Hutter, M. (2019). Learning agile and dynamic motor skills for legged robots. Science Robotics 4(26). https://arxiv.org/pdf/1901.08652.pdf

[4] Lidec, Q. L., Montaut, L., Schmid, C., Laptev, I., & Carpentier, J. (2022). Leveraging randomized smoothing for optimal control of nonsmooth dynamical systems. arXiv preprint arXiv:2203.03986. https://arxiv.org/pdf/2203.03986

[5] Suh, H. J., Simchowitz, M., Zhang, K., & Tedrake, R. (2022, June). Do differentiable simulators give better policy gradients?. In International Conference on Machine Learning (pp. 20668-20696). PMLR. https://proceedings.mlr.press/v162/suh22b/suh22b.pdf

[6] Mordatch, I., Lowrey, K., Andrew, G., Popovic, Z., and Todorov, E. (2015). Interactive Control of Diverse Complex Characters with Neural Networks. Advances in neural information processing systems, 28. https://proceedings.neurips.cc/paper_files/paper/2015/file/2612aa892d962d6f8056b195ca6e550d-Paper.pdf

[7] Hansen, N. (2016). The CMA evolution strategy: A tutorial. arXiv preprint arXiv:1604.00772. https://arxiv.org/pdf/1604.00772.pdf

[8] Toussaint, M. (2017). A tutorial on Newton methods for constrained trajectory optimization and relations to SLAM, Gaussian Process smoothing, optimal control, and probabilistic inference. Geometric and numerical foundations of movements, 361-392. https://argmin.lis.tu-berlin.de/papers/17-toussaint-Newton.pdf

[9] Rajamäki, J., Naderi, K., Kyrki, V., & Hämäläinen, P. (2016, October). Sampled differential dynamic programming. In IEEE/RSJ IROS. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7759229

[10] Hämäläinen, P., Rajamäki, J., & Liu, C. K. (2015). Online control of simulated humanoids using particle belief propagation. ACM Transactions on Graphics (TOG), 34(4), 1-13. https://dl.acm.org/doi/pdf/10.1145/2767002

[11] Guo, B., Jiang, Y., Kamgarpour, M., & Ferrari-Trecate, G. (2023, June). Safe zeroth-order convex optimization using quadratic local approximations. In IEEE CDC 2023. https://arxiv.org/pdf/2211.02645.pdf

[12] Manchester, Z., & Kuindersma, S. (2016, December). Derivative-free trajectory optimization with unscented dynamic programming. In 2016 IEEE 55th Conference on Decision and Control (CDC) (pp. 3642-3647). https://roboticexplorationlab.org/papers/udp.pdf

[13] Ortiz, J., Pupilli, M., Leutenegger, S., & Davison, A. J. (2020). Bundle adjustment on a graph processor. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2416-2425). https://openaccess.thecvf.com/content_CVPR_2020/papers/Ortiz_Bundle_Adjustment_on_a_Graph_Processor_CVPR_2020_paper.pdf

[14] Levine, S., and Koltun, V. (2013) Guided policy search. International conference on machine learning. http://proceedings.mlr.press/v28/levine13.pdf