Paper Mini-Critical Review: Interdisciplinary Research Methods

Chirayu Salgarkar

October 24, 2024

Problem #1. Title: Safe exploration in model-based reinforcement learning using control barrier functions

Problem #2. Authors: Max H. Cohen, Calin Belta, of Boston Univ.

Problem #3. The paper provides contributions in the fields of safe control and implement it into Reinforcement Learning paradigms They begin by creating a new class of CBF. These CBF's, denoted Lyapunov-like Control Barrier Functions (LCBFs), have one key distinction to standard CBFs: the barrier function vanishes at the origin, and they are always positive semi-definite along the set-interior. these conditions render the set-interior forward invariant and therefore safe. This LCBF can be used to construct a controller, which can be applied onto a Reinforcement Learning Policy, while maintaining safety. They then utilize this formulation for development of a model-based reinforcement learning framework (MBRL) for online learning of the value function of an optimal control problem, while implementing control barrier functions (CBFs) to guarantee the obeying of safety constraints. While the MBRL paradigm is inspired by Kamalapurkar et al. (Kamalapurkar, Rosenfeld, & Dixon, 2016), their contribution presents a method for learning a performance-driven policy safely, which enables less conservative policy generation. This paper then provides numerical example validation for these claims and demonstrate that this method allows for stabilization and obstacle avoidance concurrently. Prior to this work, research was unclear regarding how to achieve safety in non-convex constraints, or in environments where RL policies, which are often performance-driven, could violate safety without further intervention. This approach provides a method for generating such safe policy while also allowing for exploration.

Problem #4. Below are the two papers I have used for this section.

0.1 Paper 1

My first paper is (Vamvoudakis, Miranda, & Hespanha, 2016). In this paper, the authors propose a control algorithm to solve the *infinite-horizon optimal control problem*, or the pollicy that maximizes reward over an indefinite time period for nonlinear systems with actuator saturation. To do so, the authors use actor-critic policy, where the critic learns the optimal cost while the actor learns the optimal policy, and then incorporating a separate robustness term onto the controller. This method allowed for asymptotic stability of the closed-loop. This paper is fundamental for understanding the work at hand. Both attempt to solve the infinite horizon optimal control problem, using adaptive dynamic programming and RL to solve optimal control problems. If anything this paper can be seen as a seminal parent paper on the topic. More recent work, including the current paper being critiqued is based off this earlier work. While this work solves unconstrained Optimal Control, more work now is interested in the incorporation of safety constraints while doing so.

0.2 Paper 2

My second paper is (Marvi & Kiumarsi, 2021). This article proposes another safe RL paradigm, but this time through the augmentation of a CBF candidate onto the cost function. The RL algorithm learns the optimal control policy to minimize this cost function, which then gets implemented onto a controller, ultimately guaranteeing safety onto the controller (as it forces the controller to stay in the safe set). They then test this prolicy through an automotive case of lane-keeping. Marvi and Kiumarsi's techniques require the assumption that the value function is continuously differentiable, which may not be true. It is this reason that motivates Cohen and Belta to *decouple* learning and safety, making them independent, so that you can safely learn a CBF or alternative while dealing with potentially partially uncertain dynamics. While both papers incorporate CBFs into RL, they do so differently, one by integrating CBFs into an off-policy framework, while the other incorporates LCBFs into a model-based setting, prioritising independent learning and safety.

Problem #5. As in many papers in this field, their methodology begins with a *problem formulation*. Their problem formulation deemed to find a control policy ensuring safety, which is measurable by forward invariance (keeping the state $x(t)$ inside a safe set C at all times) while simultaneously learning a value function and optimal policy to minimize cost over an infinite time horizon. Their methodology continues in their mathematical analysis, by developing the LCBFs. They then conduct mathematical proofs to demonstrate that these LCBFs guarantee conditions like forward invariance. This setup of mathematical problem formulation, development of a mathematical construct, and mathematical proof that this construct works to solve the problem formulation is consistent with methodologies in the field of control theory. The methodology then incorporates LCBFs into the overall MBRL framework and then shows mathematically that the controller still guarantees safety constraint synthesis. Next, validation is done through numerical examples, or experimental design testing. They simulated a simple integrator system with collision/obstacle avoidance, controlled by a linear quadratic regulation policy as weell as a learning-based RL policy. The RL policy was able to successfully navigate past obstacles far more successfully than the LQR policy, and demonstrated clear system learning of dynamics and overall safety constraint validity.

Problem #6. The major future direction extended in the paper is the incorporation of *Zeroing CBFs*, which explicitly guarantees convergence toward the interior of a set a time passes, which guarantees moving away from boundaries, which is very useful in collision avoidance. This is especially useful for guaranteeing forward invariance in systems with high uncertainity. Notably, zeroing CBFs would prevent cases where control inputs may have magnitudes higher than actual actuator limits, making them impossible to implement in the real world. Further research may also deal with higher uncertainty measures, as current work only deals with the case of partial knowledge of drift dynamics. While this is a substantial improvement to the current literature, future work could deal with handling more severe uncertainties through other RL paradigms, such as model-free controllers. Further research could also be in the implementation side, such as applying such a framework directly onto autonomous vehicles to measure real-world robustness.

Problem #7. While this paper is relatively strong, there are still some directions that are not considered in this paper. As stated earlier, this current work uses Model-Based RL policies, with a justification that Model-free methods, which implement safety and policy generation concurrently, would be at both. However, this is still not good enough in situations where you do not have any knowledge of system dynamics, the motivation behind Policy Gradient and other methods to begin with. It is necessary to develop safe offline RL policies, and the implementation of CBFs may still be possible through this. Secondly, this problem could be investigated in a transfer learning context - that is, give a safe control poicy from one environment and test it in a similar but new environment, and see how successful the control policy is, especially when learning safety hyperparameters, which take a lot of computational time. This may reduce load times and allow for better real-world implementation. This would also be interesting to explore from a stochastic perspective, studying the strength of the CBF-based controller with random disturbances present. Finally, I would be interested in incorporating other methods of guaranteeing safety, like Hamilton-Jacobi reachability or reachable tubes, and compare performance to the previous paradigm. This could be accomplished by a mathematical formulation of those tubes into the set and then comparing it with the method above.

References

Kamalapurkar, R., Rosenfeld, J. A., & Dixon, W. E. (2016). Efficient model-based reinforcement learning for approximate online optimal control. *Automatica*, *74*, 247-258. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0005109816303272` doi: https://doi.org/10.1016/j.automatica.2016.08.004

Marvi, Z., & Kiumarsi, B. (2021). Safe reinforcement learning: A control barrier function optimization approach. *International Journal of Robust and Nonlinear Control*, *31*(6), 1923–1940.

Vamvoudakis, K. G., Miranda, M. F., & Hespanha, J. P. (2016). Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(11), 2386-2398. doi: 10.1109/TNNLS.2015.2487972