Viability in State-Action Space: Connecting Mechanics, Control and Learning.

Steve Heim

September 2019

Abstract

Lorem Ipsum Doloret.

Zusammenfassung

Ackenschmackenfracken.

Chapter 1

Introduction

Mobility is a driver of revolutions. Needs to be fast, cheap and reliable. Robustness is always a key for this.

Chapter 2

Preliminaries

We introduce here relevant background in a concise manner.

2.1 Reinforcement Learning

Reinforcement learning is cool. Introduce MDPs. Some notes on terminology... cost is reward.

2.1.1 Model-free and model-based

RL is just DPA, done model free. It has these advantages. It is basically trying to learn to negotiate the reward-landscape. It is, however, inefficient and does not extrapolate well. Models should be used when they are good.

2.1.2 Shaping

In RL, shaping refers to shaping the reward-landscape, in order to make it easier to for a learning agent to learn on. Most of the effort in this field is called curriculum design: the agent learns on easier tasks which inform it about the original policy. There can also be shaping by modifying the reward function. This can be a temporary, hand-crafted change, or it can be IRL. It is akin to solving a convex approximation of the cost-function in model-based optimization. As RL often takes a model-free approach, usually there is no emphasis on lower and upper bounds (as there is in model-based optimization). Mostly focused on changing R in the MDP, in some cases a combination of R and P.

2.2 Viability and Backreachability

Introduce viability and backreachability.

Chapter 3

Shaping with Training Wheels

Most shaping is done by focusing on R. There has been, to the best of our knowledge, only one study which formally explores the effect of changing P, isolated from changes in R. It is however, purely done in simulation. We explore this concept empirically for a hopper.

3.0.1 Concept

We have a hopper, with a parameterized oscillatory controller. We perform stochastic gradient descent to find controller parameters. Due to the low dimensionality, we can empirically map out the entire reward landscape quickly and reliably.

3.0.2 Choice of Training Wheel

3.0.3 Shallower Gradients

3.0.4 Salient Gradient Sets

3.0.5 Conclusion

Chapter 4

Theory of Viability in State Action Space

We introduce in this chapter the formal mathematics of extending the notion of viability theory to state-action space. We will cover

4.1 Viable Sets in State Action Space

Formal Definition, and an algorithm to compute based on the transition matrix Conditions for a policy to be viable.

4.2 Measures of Viability

Introduce the measure, taken over the viable set, and of the whole set, or of a slice at a point of the state-space. Map this back into Q-space.

4.3 Robustness to Uncertainty

We can now formalize robustness to uncertainty in action space. This can be considered as a minimax or adversarial. Point out relation to uncertainty in state space.

Chapter 5

Applications of Viability in State Action Space

We will do the following:

5.1 Safe Model Free Learning

5.2 Quantify Robustness of Natural Dynamics

Since any viable policy has to live in the viable set, we can compare systems with different natural dynamics. Define Natural Dynamics.

5.2.1 Hierarchical Control, and templates and anchors

We consider the natural dynamics of the underlying ”blackbox”.

5.2.2 SLIP and NSLIP

Connection to bifurcation analysis, and why this shows so much more. In state-space, viability kernel is exactly the same.

5.2.3 Blackbox optimization of robustness

Not prescriptive. Conclusion: slow (relies on brute force), but still useful for analysis, and because design can be done offline and it is embarrassingly parallel. Would benefit from improving scalability, and from approximations.

5.3 Safe, Model-free Learning

We use this to formalize safe model-free learning. Collaboration with Alexander von Rohr and Sebastian Trimpe, who in particular brought in expertise in Gaussian processes and active sampling.

5.3.1 Related work

5.3.2 Using the Measure as a Safety Function

5.3.3 Modeling the Measure as a Gaussian Process

5.3.4 Learning the Measure by Sampling

5.3.5 Results

5.3.6 Discussion and Outlook

5.4 Learning from outside the Viability Kernel

We have seen from shaping that the reward landscape depends on all factors in the MDP.

Chapter 6

Discussion and Outlook