

ESE 450 Design Document Draft 1

Alexandre Amice, Sophia Moses, Sergio Roman, Matteo Sciolla

September 5, 2019

Abstract

Training a policy controller, whether with a human or through machine learning techniques is often a costly process. However, once this effort is expended, a good policy can effectively control a dynamic system from many states. However, in cases where the certain actuators in the system break, the original policy of the intelligent agent is no longer an effective strategy for controlling the system. In this project, we seek to design an intermediate controller which can map the actions of the original policy to a new set of actions which will control the damaged dynamic system to the same objective state as the original actions would under the original dynamics.

1 Control by a rational agent with a good model

Consider the dynamic system and observation model which evolve according to

$$x_{k+1} = f(x_k, u_k) \tag{1}$$

$$y_k = g(x_k) \tag{2}$$

Denote by \mathcal{S} the set of possible states x_k can assume, by \mathcal{A} the set of possible actions which can be input into the system, and by \mathcal{O} the set of possible observations.

We consider the case of rational agent **A** who attempts to control the dynamic system in (1) according to a control policy $\pi : \mathcal{O} \rightarrow \mathcal{A}$ towards some desired goal region $\mathcal{G} \subseteq \mathcal{S}$. Drawing inspiration from the work of [1], we assume that the agent **A** maintains approximation functions \tilde{f} and \tilde{g} such that for $x_k \in \mathcal{T} \subseteq \mathcal{S}$ and $u_k \in \mathcal{P} \subseteq \mathcal{A}$, then

$$\begin{aligned} \left| \tilde{f}(x_k, u_k) - f(x_k, u_k) \right| &\leq \varepsilon_f \\ |\tilde{g}(x_k) - g(x_k)| &\leq \varepsilon_g \end{aligned}$$

We call the sets \mathcal{T} and \mathcal{P} the nominal operating domains for the agent **A**. The smaller ε_f and ε_g , the better the approximation of the model. In other words, we assume that the agent **A** has a strong approximation to the dynamic system and observation model given in (1) and (2) within some bounded subregion of the state and action space. Therefore, we assume that the control policy π is well suited for achieving the goal of the agent.

2 A sudden variation in dynamics

Consider a case where in the dynamics of the system in (1) changes suddenly. In this we consider in particular the case where in certain actuators fail to perform correctly such as a single rotor in a quadrotor system suddenly jamming. Therefore, the dynamics of the system are no longer given by (1), but rather by:

$$x_{k+1} = F(x_k, u_k) \tag{3}$$

We consider the same rational agent **A** introduced in the previous section with its control policy π and its dynamic and observation approximation functions \tilde{f} and \tilde{g} attempting to control the dynamic system given in (3). Due to the failure in actuation, it is no longer possible in general to assume that in the bounded regions \mathcal{T} and \mathcal{P} , $\left| \tilde{f}(x_k, u_k) - F(x_k, u_k) \right| \leq \varepsilon_f$.

3 Helping the agent maintain control

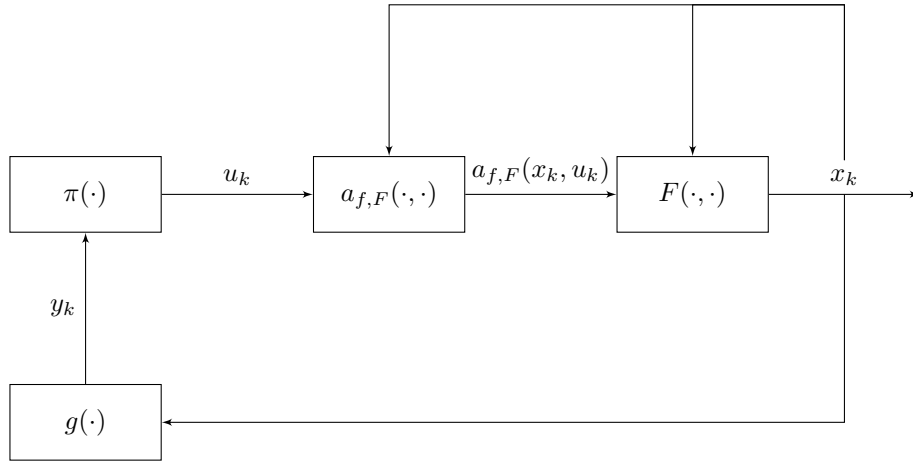
For an intelligent agent **A**, say a human or a powerful neural network, then it is typically possible to relearn a policy say π' in order to control the new dynamic system given in (3). However, retraining is expensive and costly. Additionally, when the dynamics change due to issues in actuation, the system is often in a state that still needs to be controlled in spite of the lost actuation at least until it can come to a safe stopping point. In order to circumvent this retraining issue, we propose the following scheme leveraging the rationality of the agent as well as the slow speed in with which the agent will adjust its control input.

Suppose that the failure in the system is such that there exists some function $a_{f,F} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{A}$, called an adjustment function, such that:

$$|f(x_k, u_k) - F(x_k, a_{f,F}(x_k, u_k))| \leq \varepsilon_a$$

for some states $x_k \in \mathcal{S}_a \subseteq \mathcal{S}$ and actions $u_k \in \mathcal{A}_a \subseteq \mathcal{A}$. Simply put, there exists some map $a_{f,F}$ which can remap inputs u_k such that the output dynamics of the system are the same for some inputs and in some regions of the state space. We call the region $\mathcal{S}_a \times \mathcal{A}_a$ the adjustable region.

Under this assumption, the agent **A** would have no need to change its control policy π in the region $\mathcal{S}_a \times \mathcal{A}_a$. We therefore propose the following feedback model within the adjustable region



4 Senior Design Problem

We propose to use the above framework to help a human pilot fly a drone between two points. Suppose a human has learned to control a quadrotor robot using a two joystick controller which can control the roll, pitch, yaw, and acceleration of the drone. The human is asked to fly the drone between two waypoints in space. Since the human is a rational agent who knows how to fly the drone, the human can easily maneuver the drone between the two way points.

Now suppose the human encounters a situation where one of the four rotors is jammed completely and cannot turn. The drone has lost acutation in the form of this rotor. We seek to design an adjustment function a which allows the human to continue to fly the drone with minimal retraining by mapping the human's input decisions made under the control policy of the original dynamics of the system to new inputs mapped through a which properly control the dynamics of the broken system.

References

- [1] S. Reddy, A. D. Dragan, and S. Levine, "Where do you think you're going?: Inferring beliefs about dynamics from behavior," *CoRR*, vol. abs/1805.08010, 2018.