Master Project

On the Road from Active Inference to Regret Minimization

Jesse Coehoorn, MSc student S&C and ME-BMD, TU Delft
J.D.Coehoorn@student.tudelft.nl

Martijn Wisse, Cognitive Robotics, TU Delft
M.Wisse@tudelft.nl

Peyman Mohajerin Esfahani, Delft Center for Systems and Control, TU Delft
P.MohajerinEsfahani@tudelft.nl

Context

For some time now, the Free Energy Principle, as first introduced by the neuroscientist Karl Friston, has been around to explain human behaviour and, essentially, life. By way of Active Inference, it provides a mathematical formulation of perceptual inference, learning, action and behaviour. More recently, this view on the brain has found its way into robotics and control. Roboticists are excited about the promises that Active Inference offers, but the stability properties are of yet only little understood.

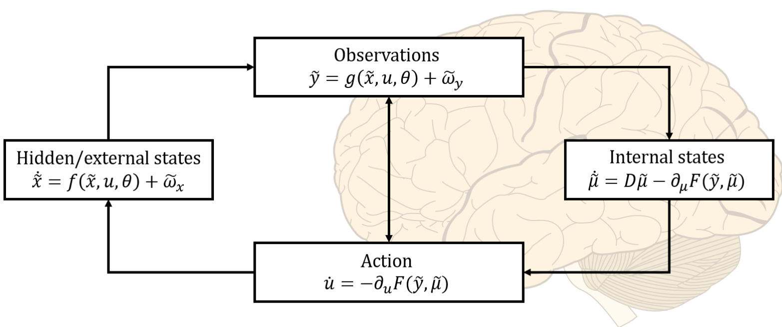

Figure 1: Schematic of Active Inference, F denotes the Free Energy

The Free Energy Principle revolves around the minimization of the Free Energy of a system by adjusting the agent's actions and internal beliefs about states, resulting in a form of prediction error minimization. This minimization is done via a gradient descent scheme. For the internal beliefs, this takes the form of a generalised descent as the beliefs also encode motion or trajectory. In the control literature, there exist more examples of control strategies that minimize some cost function, take for example the LQR setting which has been around since the sixties or the much more recent development of regret minimization in control.

Regret has been used in the Online Optimization community for decades to compare the performance of an algorithm to the best possible algorithm in hindsight. The goal is usually to make the static regret grow sublinearly, such that the average regret tends to zero when time progresses, meaning that the best actions are chosen. In a control context, the cost functions no longer depend on the action that is currently taken, but also on previous actions, leading to a notion known as policy regret. Generally, a gradient descent is employed in Online Optimization to minimize the cost functions. This warrants the search for connections between Active Inference and Online Optimization, especially between the generalised gradient descent and the gradient descent used for policy regret, as there might be some similarities between encoded motion and policy regret.

Furthermore, this project tries to come up with stability guarantees for Active Inference in two ways. First, a closer look is taken at how certain parameters can be adjusted to achieve closed-loop stability. The prior belief of the agent on how the system is going to evolve (the generative model) can be freely designed, just like the learning rates of the various gradient descents. The second route that will be taken is by relating Active Inference to existing solutions to cost function minimization in control. Just like the cost function in the LQR problem, the Free Energy is quadratic, and just like regret minimization, a gradient descent is executed in real-time.

Project tasks

This master thesis aims at establishing stability properties for Active Inference and exposing connections between generalised coordinates and policy regret. Before moving on to more complex systems, linear systems will be considered. Stability properties can be found via each of the following tasks (exluding the first):

1. Compare the gradient descent for policy regret with the generalised gradient descent in Active Inference
2. Identify ways to design a stabilizing generative model
3. Determine stabilizing learning rates for the gradient descent
4. Relate Active Inference to LQR control (if such a relationship exists)
5. Use insights from regret minimization in control to design new and stable algorithms to minimize Free Energy