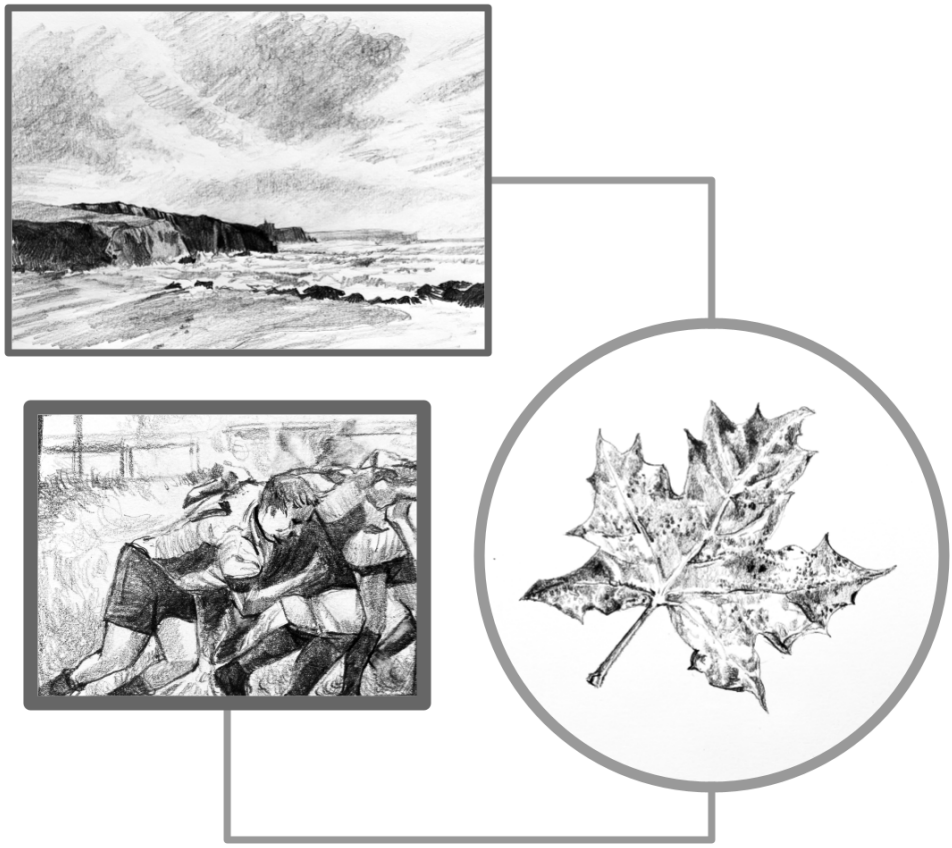

13

Optimising actions for control objectives

Concept. The idea here is

13.1 States, actions and formulating rewards

Up to this point, we have only considered actions which were either scheduled up front through some fixed process or through user interaction via a game interface. In order to start creating algorithms to act on the system state for us, we now need to develop a formalism which 'closes the loop' by feeding information back from the stochastic process to another decision-making process. Note that in most cases, the state of real-world phenomena cannot be measured perfectly. So in order to enable any agent trained on simulated phenomena to potentially act in the real world, we will need to model this measurement process as part of the information retrieval step.

Let's now define the concept of a 'measured state' $\mathcal{S}_\mathtt{t}$ of the system (some vector that doesn't have to share the same length as $X_\mathtt{t}$) at timestep $\mathtt{t}$. We can then say generally that this measured state is 'observed' using the following measurement function

$$\mathcal{S}_\mathtt{t}^i = M_\mathtt{t}^i(X', Z_\mathtt{t}, \mathtt{t})\,, \tag{13.1}$$

where we have also extended the definition of $Z_\mathtt{t}$ to include parameters which control how this measurement is performed.

In a Markov Decision Process (MDP), based on this observation alone, we would then take actions $X_{\mathtt{t}+1} = \mathcal{F}_{\mathtt{t}+1}(X', Z_\mathtt{t}, \mathcal{A}_\mathtt{t}, \mathtt{t})$, for which we would later attribute reward $\mathcal{R}_\mathtt{t}$.

Bibliography