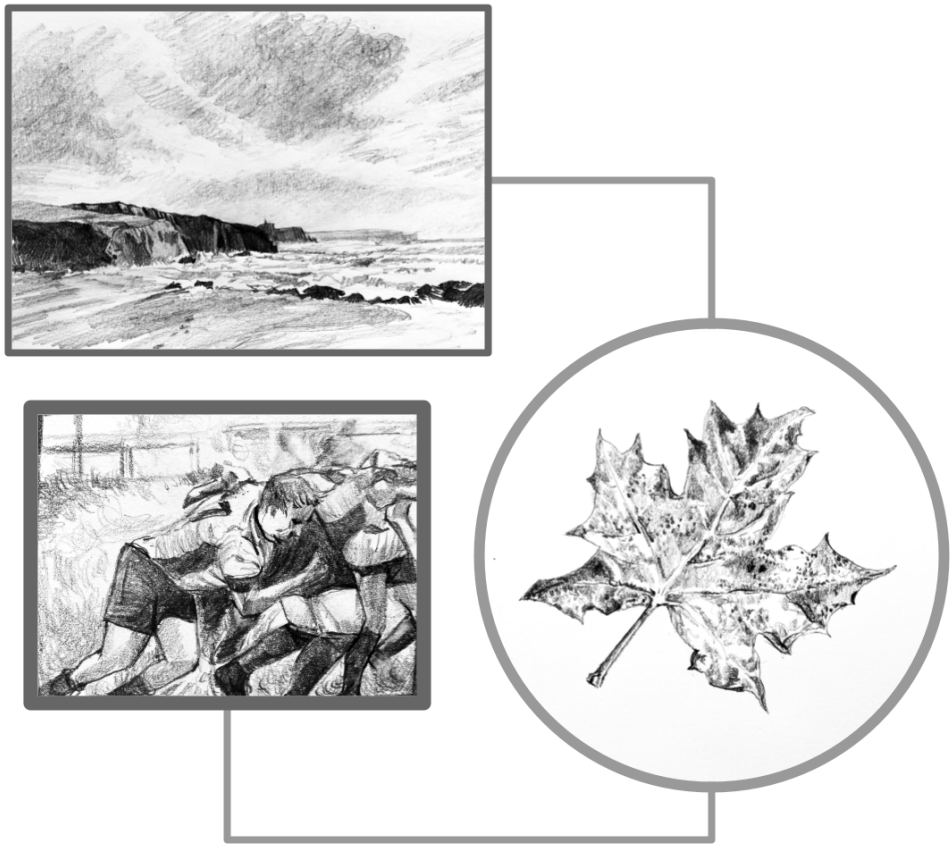

13

Optimising actions for control objectives

Concept. The idea here is

13.1 States, actions and formulating rewards

We observe the state at timestep $\mathtt{t}$ with $\mathcal{S}_\mathtt{t} = M_\mathtt{t}(X', Z_\mathtt{t}, \mathtt{t})$. 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