Why optimisation is so popular?

Optimisation and RL address different objectives:

- Optimization objective: searching an optimum a function by varying the parameters of this function
  - ➡ Optimization adapts to changes since it is usually ran from scratch
- RL maximises the cumulative reward on an MDP:
  - ➡ Runs fast if the MDP is not modified to strongly
  - ➡

RL and optimization

- All said falls into the domain of optimization:
  - ➡ An optimiser tries to find the arguments of a function to maximise the function value (optimization is greedy!)
  - ➡ RL algorithms look to find a mapping (the policy) from states to actions maximising the expected cumulative reward rather than just a single optimal function value
    - If parametric function approximation is used, we try to find the values of the parameters of the approximated function (either a value function or the policy directly) to obtain this mapping (this a classical optimisation problem).
- RL is comparable to calculus of variation (its origin is in classical mechanics - HJB equation) instead of function optimization

Optimization

$$\text{maximise}_{\{A_i\}} \sum_{t=0}^{T} R(S_t, A_t, W_t)$$

$$\text{subject to: } S_{t+1} = f(S_t, A_t, W_t)$$

RL

$$\text{maximise}_{\pi_t} \mathbb{E}_{W_t}[\sum_{t=0}^{T} R_t(S_t, A_t, W_t)]$$

$$\text{subject to: } S_{t+1} = f(S_t, A_t, W_t)$$

$$A_t = \pi_t(S_t, S_{t-1}, \ldots)$$

Feedback structure takes noise into account

Often optimization is performed only for one step horizon:

$$\text{maximise}_a R(\cdot, a, W_t)$$