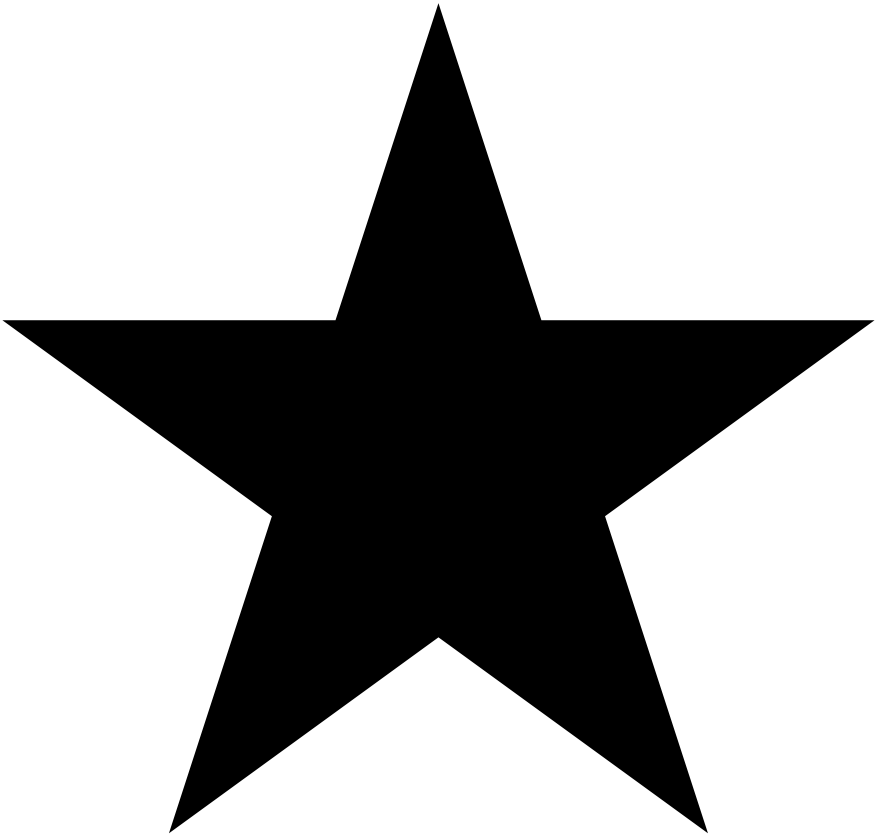
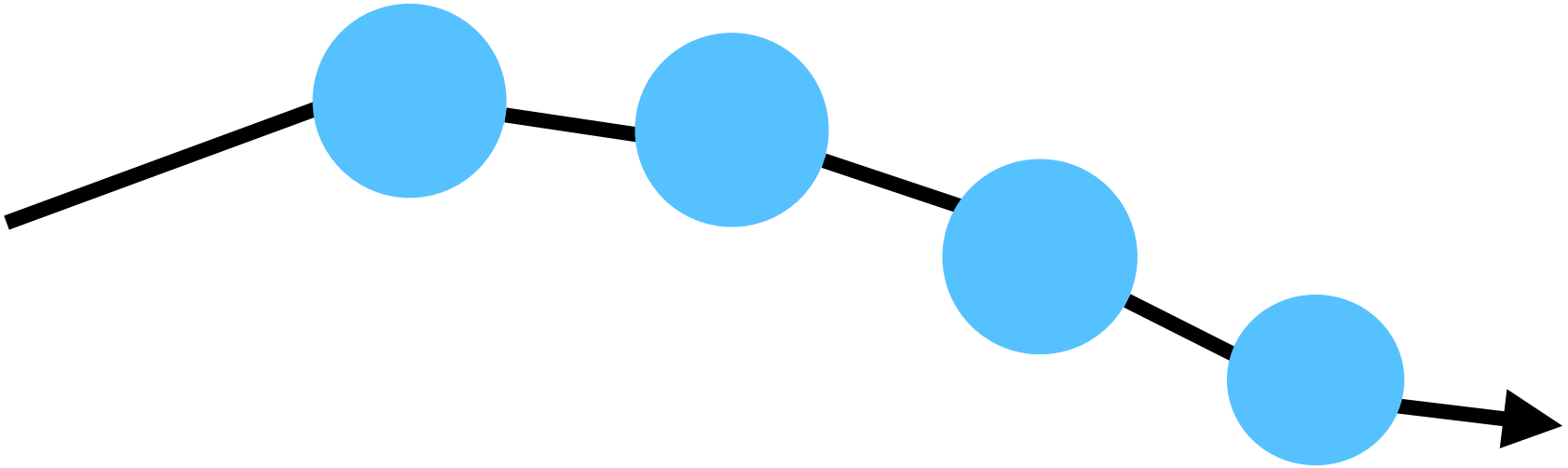
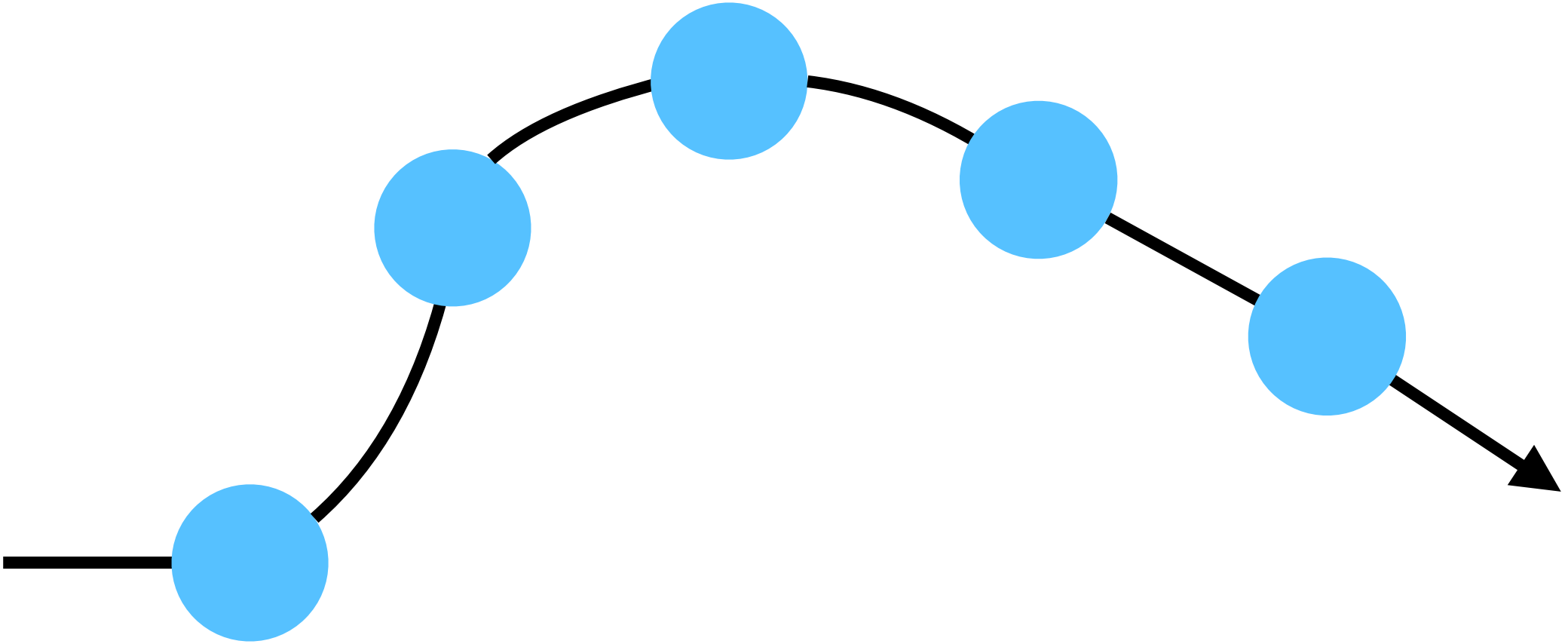
Decision-time Planning

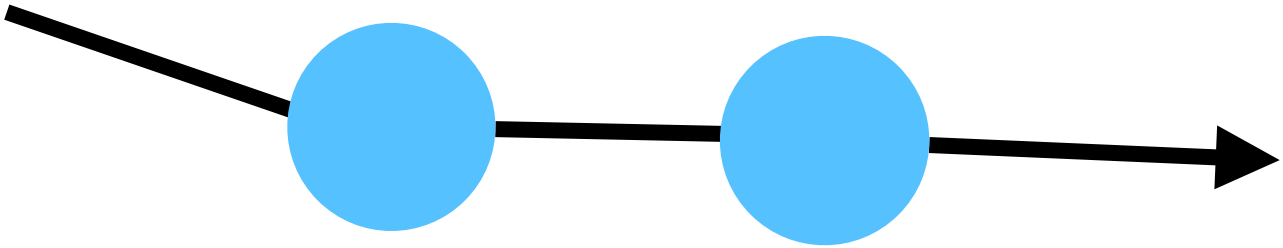
Model Predictive Control (MPC)

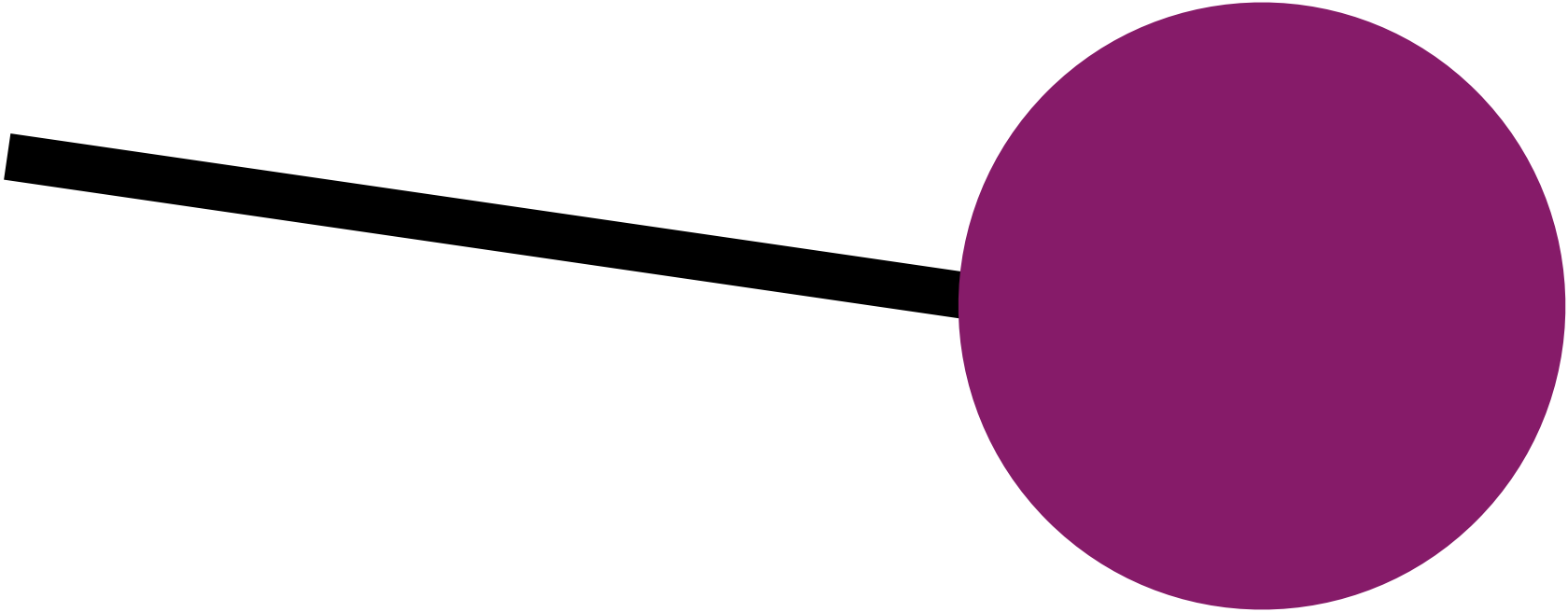
29

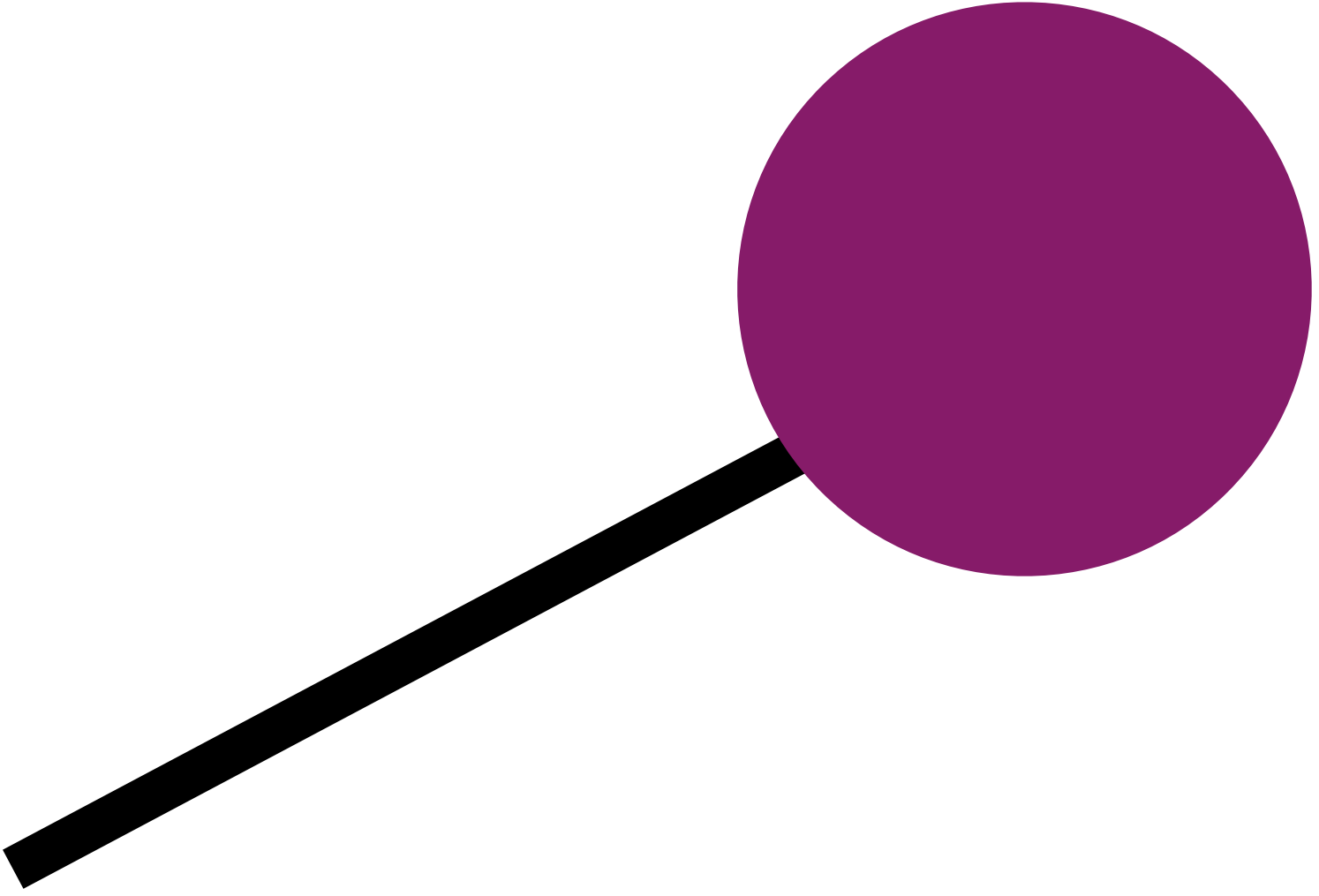


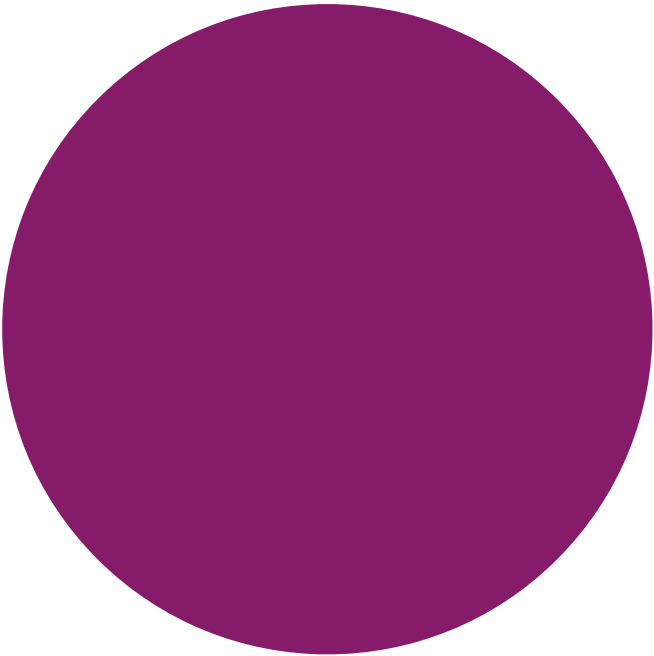












Diverged from planned trajectory…

Discard $a_1, \ldots, a_H$

So let’s replan.

And so on…

For each environment step

Observe state s

Plan $a_{0:H}$ to maximise return $\sum_{t=0}^{H-1} \gamma^t r(s_t, a_t) + \gamma^H Q_\theta(s_H, a_H)$

Execute a_0 and discard $a_1, \ldots, a_H$

Any trajectory optimisation method

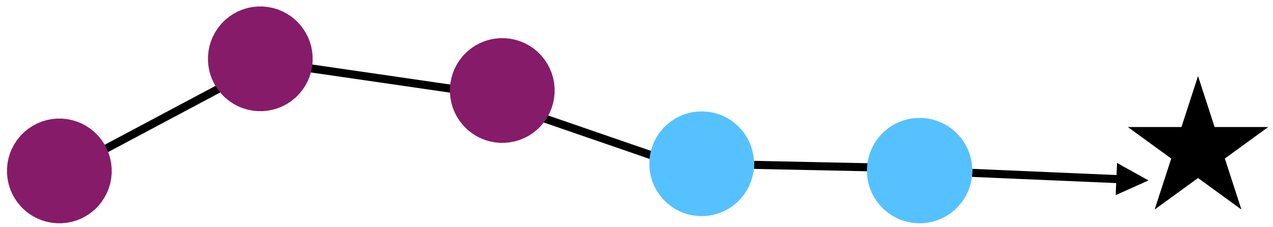
Diverged from planned trajectory…

Discard $a_1, \ldots, a_H$

So let's replan.

Decision-time Planning

Model Predictive Control (MPC)



And so on…

For each environment step

Observe state s

Any trajectory optimisation method

Plan $a_{0:H}$ to maximise return $\sum_{t=0}^{H-1} \gamma^t r(s_t, a_t) + \gamma^H Q_\theta(s_H, a_H)$

Execute a_0 and discard $a_1, \ldots, a_H$

Decision-time Planning

Model Predictive Control (MPC)

$$\pi_{\text{MPC}}(s; f, r, Q_\theta) = \arg\max_{a_0} \max_{a_1,\ldots,a_{H-1}} = \sum_{t=0}^{H-1} \gamma^t r(s_t, a_t) + \gamma^H Q_\theta(s_H, a_H) \qquad \text{s.t.} \quad s_{t+1} = f(s_t, a_t)$$

$$s_0 = s$$