Background vs Decision-time Planning

Background planning

16

Learn how to act in any situation

Optimisation variables: θ

Parameters of policy $\pi_\theta(s)$, value $Q_\theta(s, a)$, etc

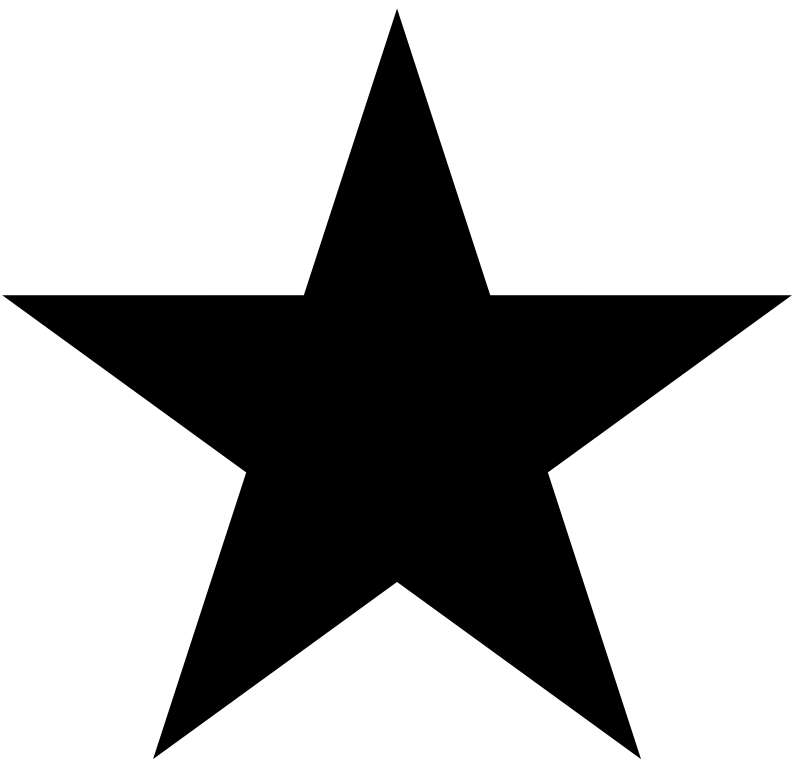
$$J(\theta) = \mathbb{E}_{s_0}\left[\sum_{t=0}^{H} r(s_t, \pi_\theta(s_t)) \right]$$

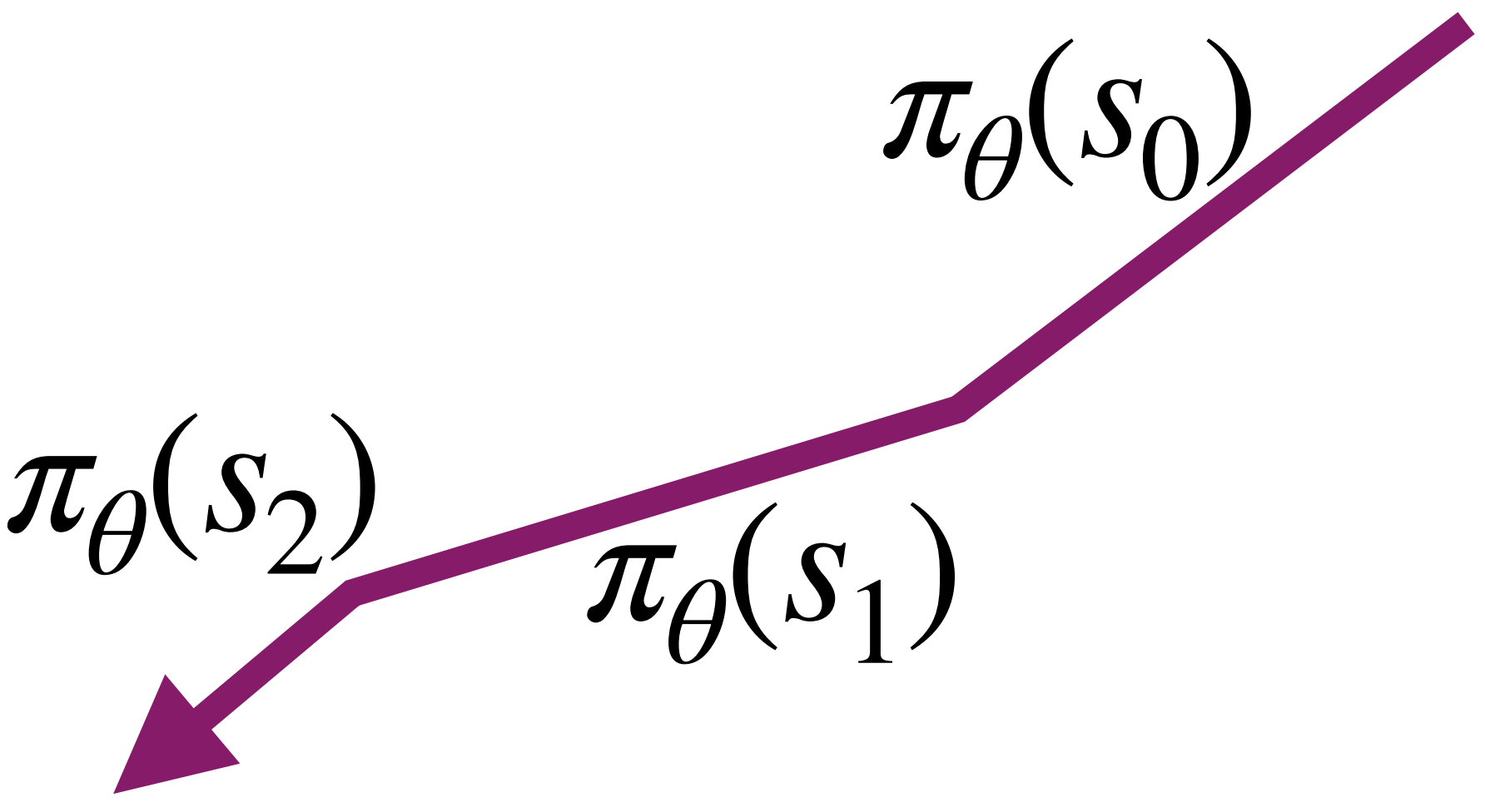
Find best action for current situation

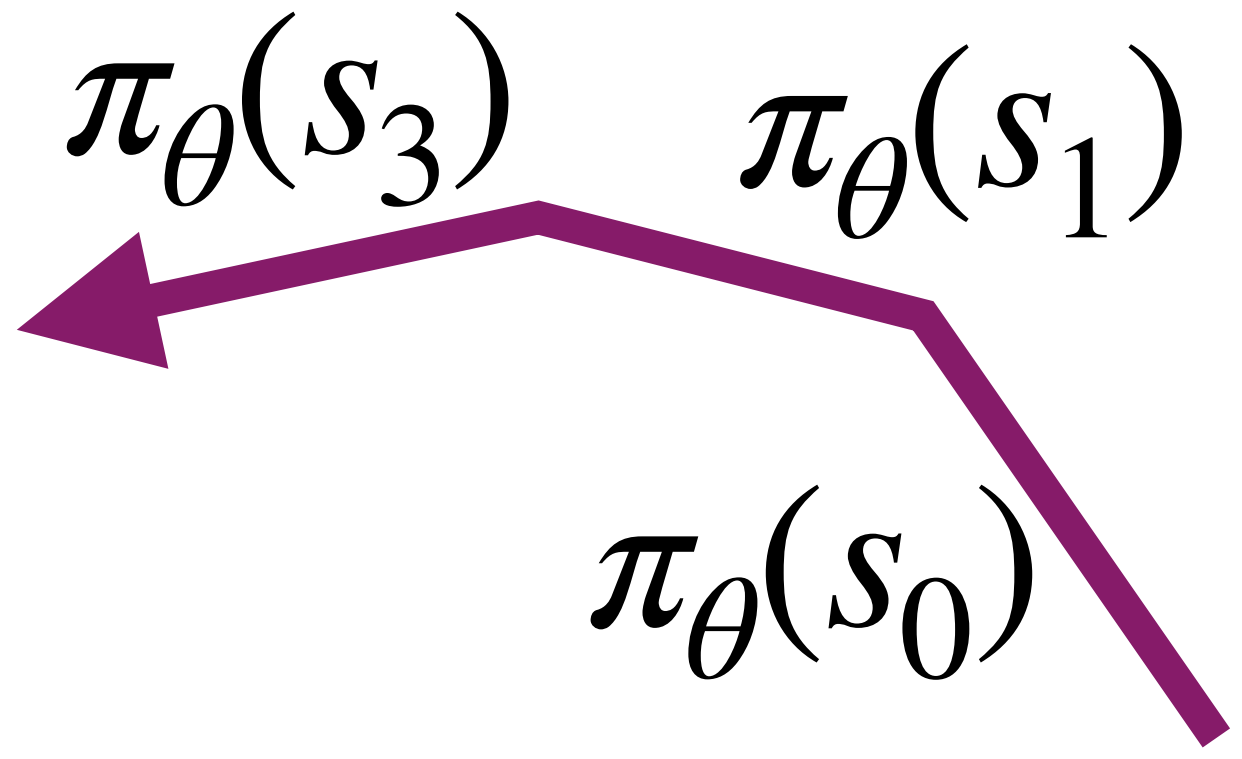
Optimisation variables: $a_0, \ldots, a_H$

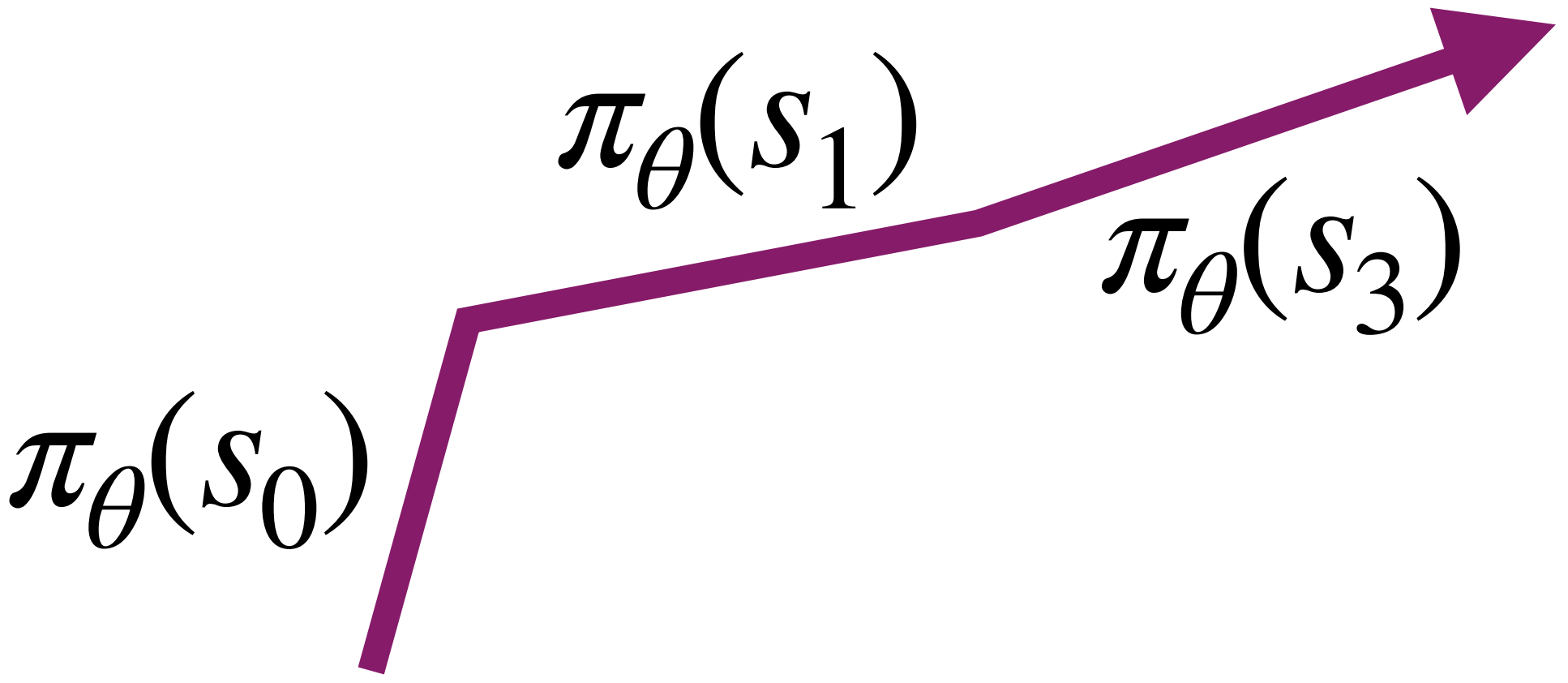
Sequence of actions (and maybe also states)

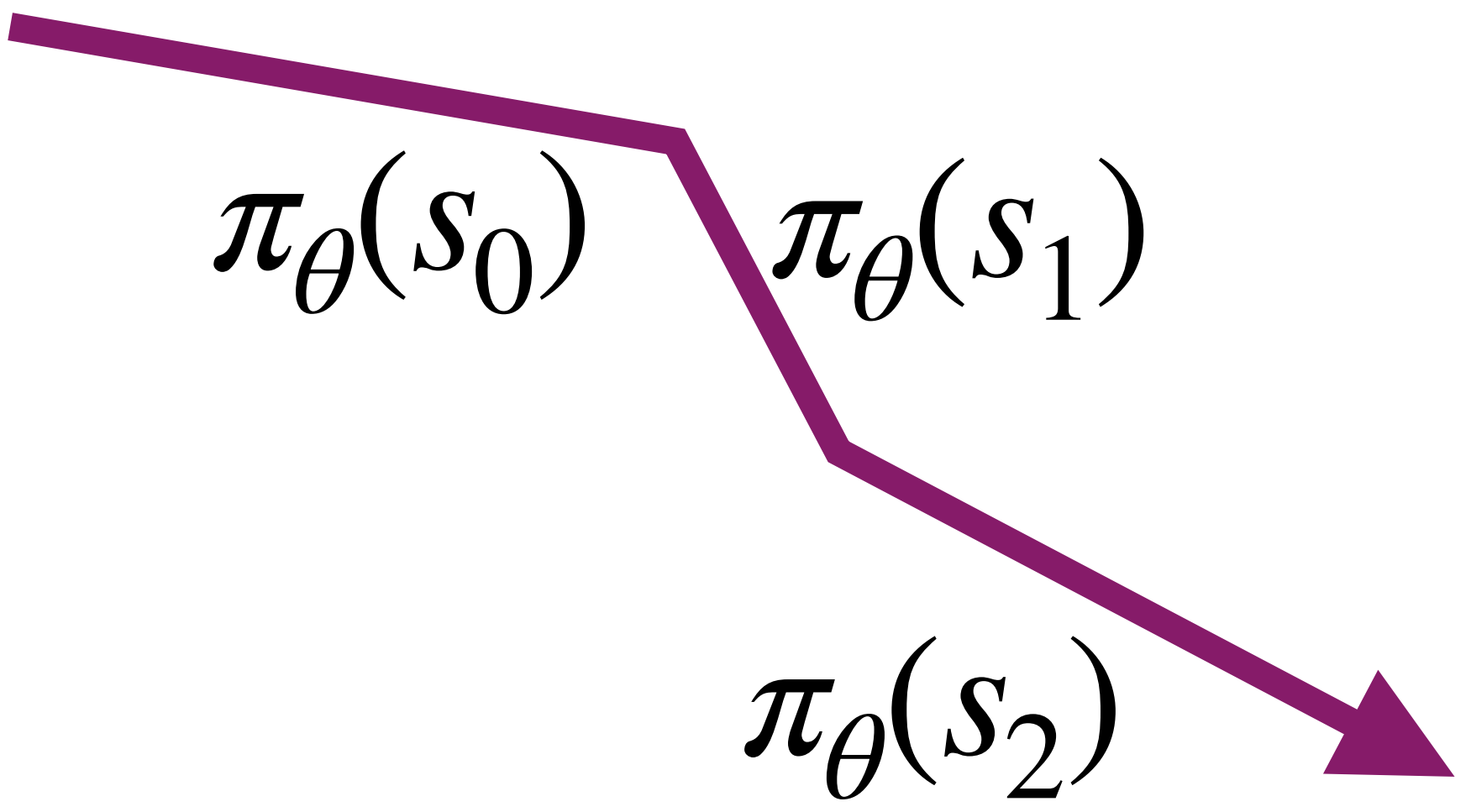
$$J(a_0, \ldots, a_H) = \sum_{t=0}^{H} r(s_t, a_t)$$

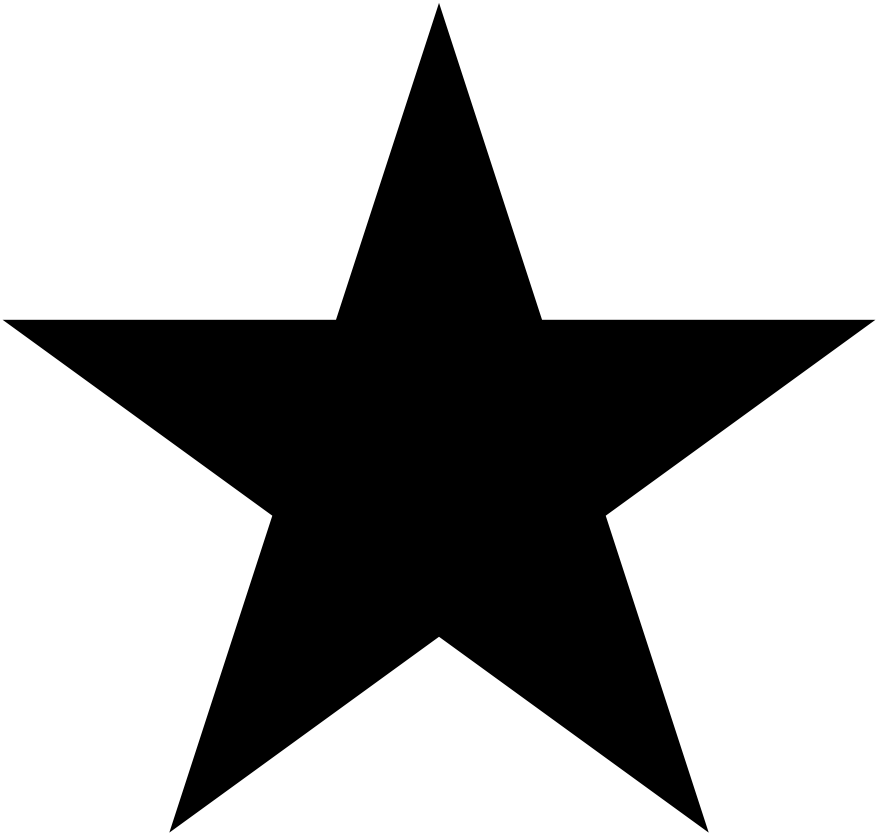



$\pi_\theta(s_0)$
$\pi_\theta(s_1)$
$\pi_\theta(s_2)$


$\pi_\theta(s_3)$
$\pi_\theta(s_1)$
$\pi_\theta(s_0)$


$\pi_\theta(s_1)$
$\pi_\theta(s_3)$
$\pi_\theta(s_0)$


$\pi_\theta(s_0)$
$\pi_\theta(s_1)$
$\pi_\theta(s_2)$




a_1
a_2
a_0

Decision-time planning

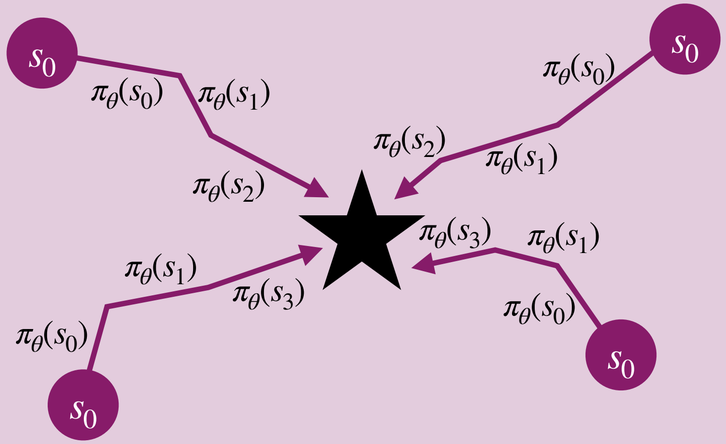

s_0
s_0
s_0
s_0


s_0

Background vs Decision-time Planning

Background planning

Learn how to act in any situation



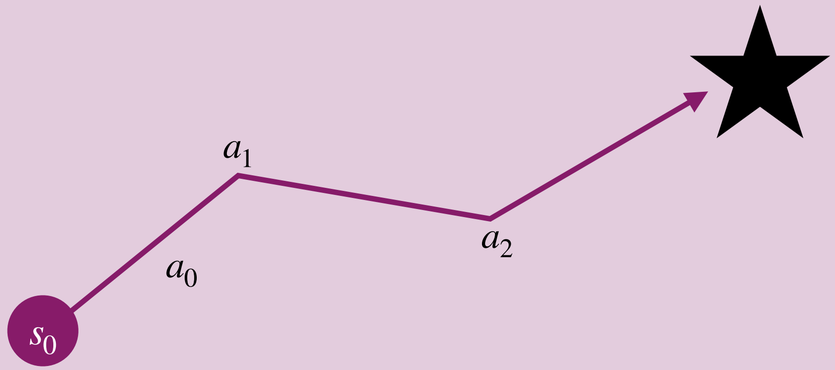
Optimisation variables: θ

Parameters of policy $\pi_\theta(s)$, value $Q_\theta(s, a)$, etc

$$J(\theta) = \mathbb{E}_{s_0}\left[\sum_{t=0}^{H} r(s_t, \pi_\theta(s_t))\right]$$

Decision-time planning

Find best action for current situation



Optimisation variables: $a_0, \ldots, a_H$

Sequence of actions (and maybe also states)

$$J(a_0, \ldots, a_H) = \sum_{t=0}^{H} r(s_t, a_t)$$

How Do We Use The "Model”?

Background vs Decision-time Planning