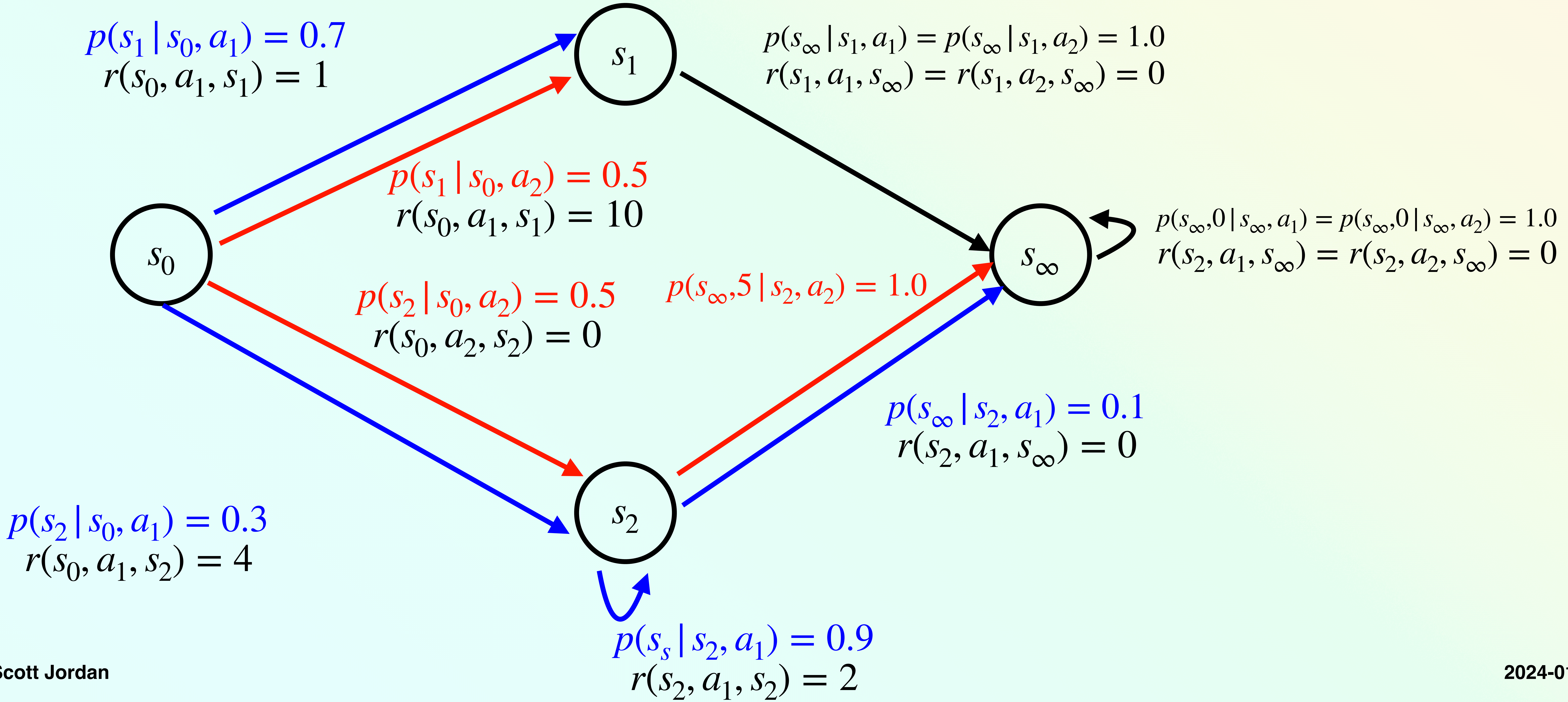
BELLMAN EQUATIONS

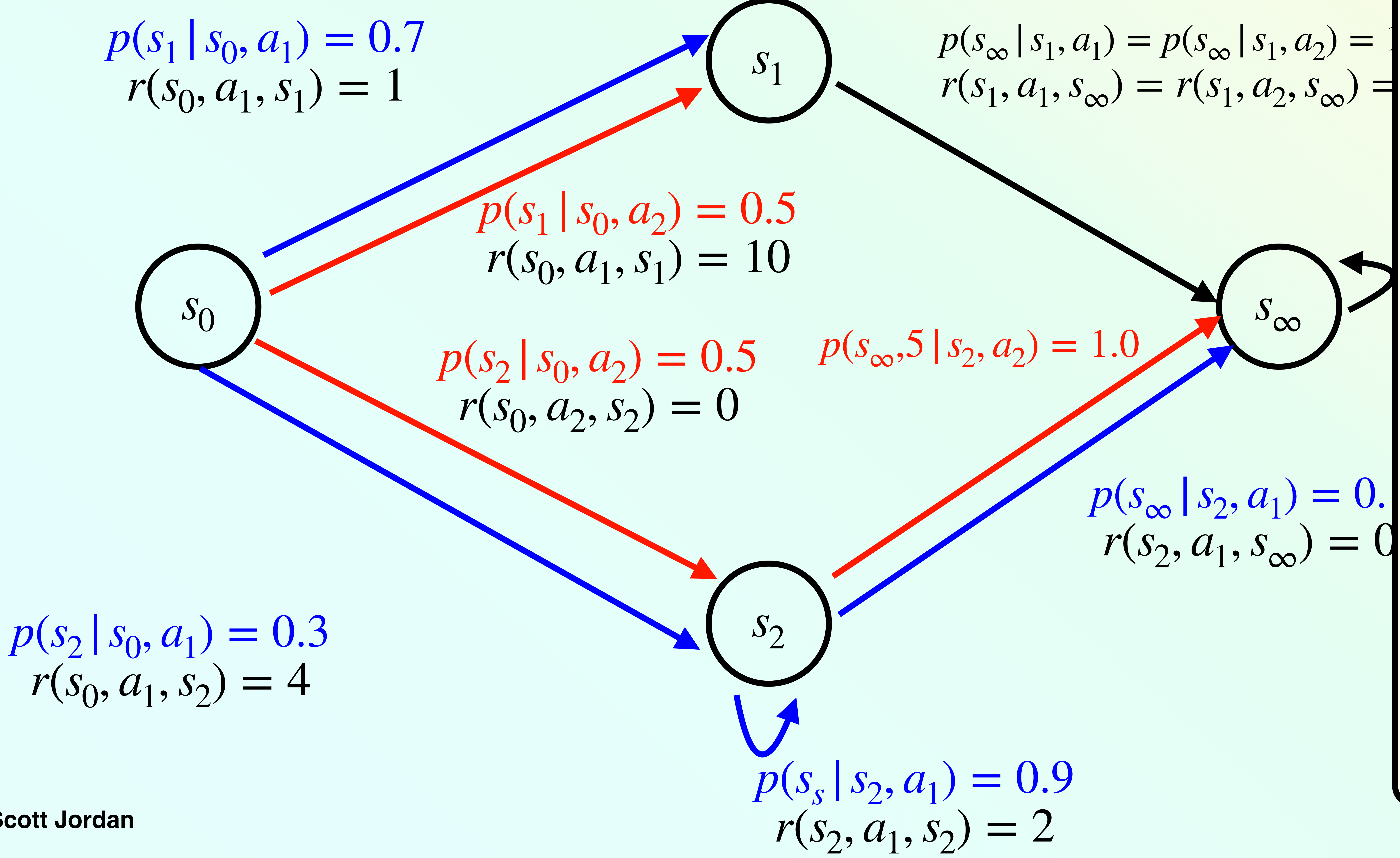
COMPUTING VALUES FUNCTIONS

EXAMPLE



COMPUTING VALUES FUNCTIONS

EXAMPLE



for all π

$$v_\pi(s_1) = 0$$

$\pi(s_2) = a_1$

$$v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$$

$\pi(s_2) = a_2$

$$v_\pi(s_2) = 5$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_1] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= r(s_0, a_1) + \gamma \sum_{s'} p(s_0, a_1, s') v_\pi(s')
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_1] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= r(s_0, a_1) + \gamma \sum_{s'} p(s_0, a_1, s') v_\pi(s') \\
&= r(s_0, a_1) + \gamma \left(p(s_0, a_1, s_1) v_\pi(s_1) + p(s_0, a_1, s_2) v_\pi(s_2) \right)
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_1] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_1] \\
&= r(s_0, a_1) + \gamma \sum_{s'} p(s_0, a_1, s') v_\pi(s') \\
&= r(s_0, a_1) + \gamma \left(p(s_0, a_1, s_1) v_\pi(s_1) + p(s_0, a_1, s_2) v_\pi(s_2) \right) \\
&= 1.9 + \gamma \left(0.7(0) + 0.3 v_\pi(s_2) \right) \\
&= 1.9 + 0.3 \gamma v_\pi(s_2)
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

For $\pi(s_2) = a_1$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= 1.9 + 0.3\gamma v_\pi(s_2) \\
&=
\end{aligned}
$$

For $\pi(s_2) = a_2$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= 1.9 + 0.3\gamma v_\pi(s_2) \\
&=
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

For $\pi(s_2) = a_1$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= \ 1.9 + 0.3\gamma v_\pi(s_2) \\
&= \ 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma}
\end{aligned}
$$

For $\pi(s_2) = a_2$

$$
\begin{aligned}
q_\pi(s_0, a_1) &= \ 1.9 + 0.3\gamma v_\pi(s_2) \\
&=
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

For $\pi(s_2) = a_1$

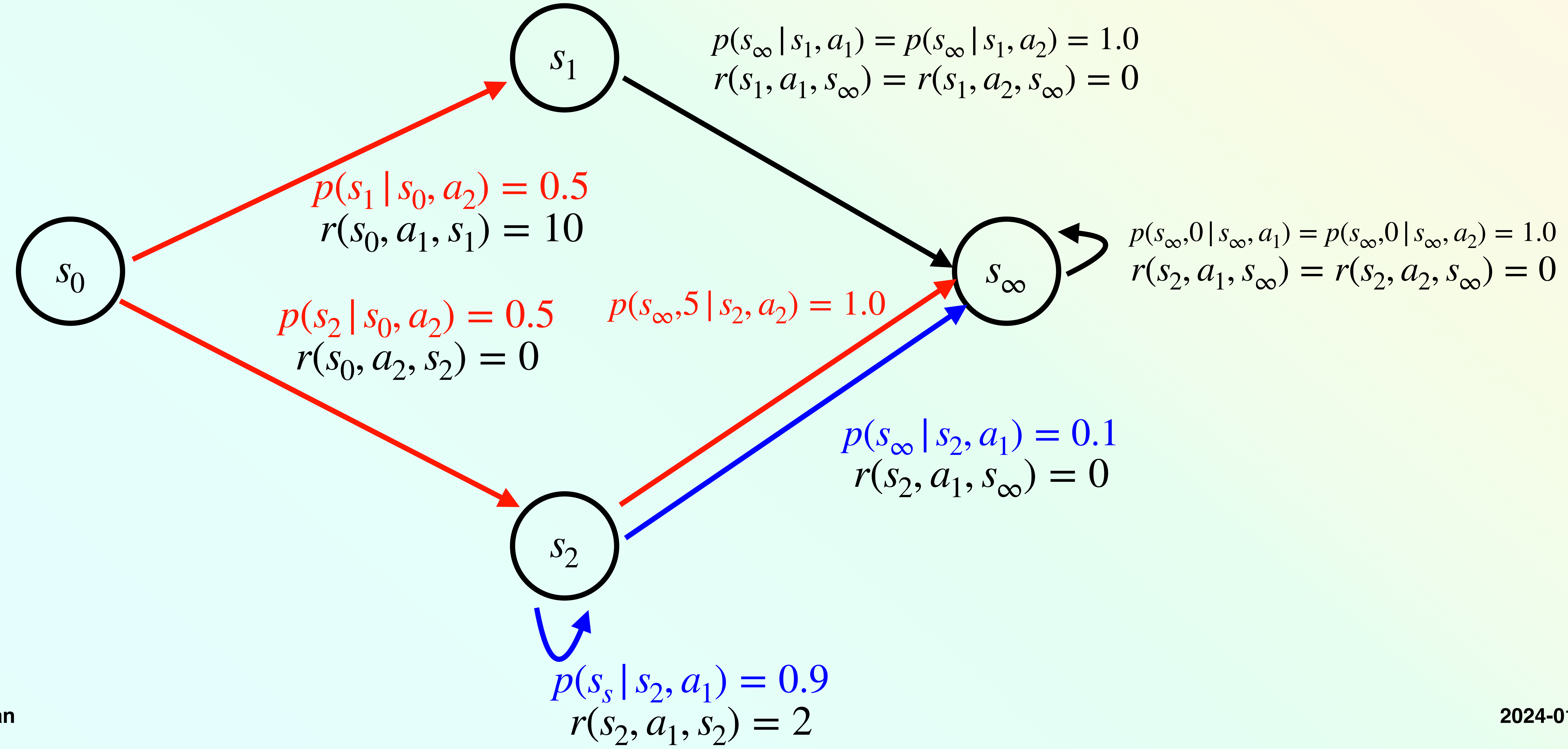
$$
\begin{aligned}
q_\pi(s_0, a_1) = &\ 1.9 + 0.3\gamma v_\pi(s_2) \\
= &\ 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma}
\end{aligned}
$$

For $\pi(s_2) = a_2$

$$
\begin{aligned}
q_\pi(s_0, a_1) = &\ 1.9 + 0.3\gamma v_\pi(s_2) \\
= &\ 1.9 + 0.3\gamma 5 = 1.9 + 1.5\gamma
\end{aligned}
$$

COMPUTING VALUES FUNCTIONS

EXAMPLE



COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_2) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= r(s_0, a_2) + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s')
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$r(s_0, a_2) = \ \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2]$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
r(s_0, a_2) &= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= \sum_{s'} p(s_0, a_2, s')\mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2, S_{t+1} = s']
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
r(s_0, a_2) &= \mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2] \\
&= \sum_{s'} p(s_0, a_2, s') \mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2, S_{t+1} = s'] \\
&= \sum_{s'} p(s_0, a_2, s') r(s_0, a_2, s')
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
r(s_0, a_2) &= \mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2] \\
&= \sum_{s'} p(s_0, a_2, s') \mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2, S_{t+1} = s'] \\
&= \sum_{s'} p(s_0, a_2, s') r(s_0, a_2, s') \\
&= p(s_0, a_2, s_1) r(s_0, a_2, s_1) + p(s_0, a_2, s_2) r(s_0, a_2, s_2)
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
r(s_0, a_2) &= \mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2] \\
&= \sum_{s'} p(s_0, a_2, s')\mathbb{E}[R_{t+1} \mid S_t = s_0, A_t = a_2, S_{t+1} = s'] \\
&= \sum_{s'} p(s_0, a_2, s') r(s_0, a_2, s') \\
&= p(s_0, a_2, s_1) r(s_0, a_2, s_1) + p(s_0, a_2, s_2) r(s_0, a_2, s_2) \\
&= 0.5(10) + 0.5(0) = 5
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_2) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= r(s_0, a_2) + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s')
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_2) &= \mathbb{E}[G_t \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= \mathbb{E}[R_{t+1} \,|\, S_t = s_0, A_t = a_2] + \gamma \mathbb{E}[G_{t+1} \,|\, S_t = s_0, A_t = a_2] \\
&= r(s_0, a_2) + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s') \\
&= 5 + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s')
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$q_\pi(s_0, a_2) = \ 5 + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s')$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$
\begin{aligned}
q_\pi(s_0, a_2) &= 5 + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s') \\
&= 5 + \gamma \left(p(s_0, a_2, s_1) v_\pi(s_1) + p(s_0, a_2, s_2) v_\pi(s_2) \right)
\end{aligned}
$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$\begin{aligned} q_\pi(s_0, a_2) &= 5 + \gamma \sum_{s'} p(s_0, a_2, s') v_\pi(s') \\ &= 5 + \gamma \left(p(s_0, a_2, s_1) v_\pi(s_1) + p(s_0, a_2, s_2) v_\pi(s_2) \right) \\ &= 5 + \gamma \left(0.5(0) + 0.5 v_\pi(s_2) \right) \end{aligned}$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$\begin{aligned}
q_\pi(s_0, a_2) &= \ 5 + \gamma \sum_{s'} p(s_0, a_2, s')v_\pi(s') \\
&= \ 5 + \gamma \left(p(s_0, a_2, s_1)v_\pi(s_1) + p(s_0, a_2, s_2)v_\pi(s_2)\right) \\
&= \ 5 + \gamma \left(0.5(0) + 0.5v_\pi(s_2)\right) \\
&= \ 5 + 0.5\gamma v_\pi(s_2)
\end{aligned}$$

COMPUTING VALUE FUNCTIONWS

$q_\pi(s_0, a)$

$$q_\pi(s_0, a_2) = \; 5 + 0.5\gamma v_\pi(s_2)$$

For $\pi(s_2) = a_1$

$$q_\pi(s_0, a_2) = \; 5 + 0.5\gamma \frac{2}{1 - 0.9\gamma}$$

For $\pi(s_2) = a_2$

$$q_\pi(s_0, a_2) = \; 5 + 2.5\gamma$$

OPTIMAL POLICY

EXAMPLE

$$\pi(s_0) = a_1, \pi(s_2) = a_1 \qquad v_\pi(s_0) = 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma} \quad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$$

OPTIMAL POLICY

EXAMPLE

$$\pi(s_0) = a_1, \pi(s_2) = a_1 \quad v_\pi(s_0) = 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma} \quad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$$

$$\pi(s_0) = a_1, \pi(s_2) = a_2 \quad v_\pi(s_0) = 1.9 + 1.5\gamma \quad v_\pi(s_2) = 5$$

OPTIMAL POLICY

EXAMPLE

$$\pi(s_0) = a_1, \pi(s_2) = a_1 \qquad v_\pi(s_0) = 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma} \qquad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$$

$$\pi(s_0) = a_1, \pi(s_2) = a_2 \qquad v_\pi(s_0) = 1.9 + 1.5\gamma \qquad v_\pi(s_2) = 5$$

$$\pi(s_0) = a_2, \pi(s_2) = a_1 \qquad v_\pi(s_0) = 5 + 0.5\gamma \frac{2}{1 - 0.9\gamma} \qquad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$$

$$\pi(s_0) = a_2, \pi(s_2) = a_2 \qquad v_\pi(s_0) = 5 + 2.5\gamma \qquad v_\pi(s_2) = 5$$

OPTIMAL POLICY

EXAMPLE: $\gamma \in [0,2/3)$

$\pi(s_0) = a_1, \pi(s_2) = a_1 \quad v_\pi(s_0) = 1.9 + 0.3\gamma\frac{2}{1-0.9\gamma} \quad v_\pi(s_2) = \frac{2}{1-0.9\gamma}$

$\pi(s_0) = a_1, \pi(s_2) = a_2 \quad v_\pi(s_0) = 1.9 + 1.5\gamma \quad v_\pi(s_2) = 5$

$\pi(s_0) = a_2, \pi(s_2) = a_1 \quad v_\pi(s_0) = 5 + 0.5\gamma\frac{2}{1-0.9\gamma} \quad v_\pi(s_2) = \frac{2}{1-0.9\gamma}$

$\pi(s_0) = a_2, \pi(s_2) = a_2 \quad v_\pi(s_0) = 5 + 2.5\gamma \quad v_\pi(s_2) = 5$

OPTIMAL POLICY

EXAMPLE: $\gamma \in (2/3,1]$

$\pi(s_0) = a_1, \pi(s_2) = a_1$ $\quad v_\pi(s_0) = 1.9 + 0.3\gamma\frac{2}{1-0.9\gamma}$ $\quad v_\pi(s_2) = \frac{2}{1-0.9\gamma}$

$\pi(s_0) = a_1, \pi(s_2) = a_2$ $\quad v_\pi(s_0) = 1.9 + 1.5\gamma$ $\quad v_\pi(s_2) = 5$

$\pi(s_0) = a_2, \pi(s_2) = a_1$ $\quad v_\pi(s_0) = 5 + 0.5\gamma\frac{2}{1-0.9\gamma}$ $\quad v_\pi(s_2) = \frac{2}{1-0.9\gamma}$

$\pi(s_0) = a_2, \pi(s_2) = a_2$ $\quad v_\pi(s_0) = 5 + 2.5\gamma$ $\quad v_\pi(s_2) = 5$

OPTIMAL POLICY

EXAMPLE: $\gamma = 2/3$

$\pi(s_0) = a_1, \pi(s_2) = a_1$ $\quad v_\pi(s_0) = 1.9 + 0.3\gamma \frac{2}{1 - 0.9\gamma}$ $\quad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$

$\pi(s_0) = a_1, \pi(s_2) = a_2$ $\quad v_\pi(s_0) = 1.9 + 1.5\gamma$ $\quad v_\pi(s_2) = 5$

$\pi(s_0) = a_2, \pi(s_2) = a_1$ $\quad v_\pi(s_0) = 5 + 0.5\gamma \frac{2}{1 - 0.9\gamma}$ $\quad v_\pi(s_2) = \frac{2}{1 - 0.9\gamma}$

$\pi(s_0) = a_2, \pi(s_2) = a_2$ $\quad v_\pi(s_0) = 5 + 2.5\gamma$ $\quad v_\pi(s_2) = 5$

BELLMAN EQUATIONS

EXPRESSING VALUE FUNCTIONS IN TERMS OF VALUE FUNCTIONS IN OTHER STATES

$$\underbrace{q_\pi(s,a)}_{\text{value}} = \underbrace{r(s,a)}_{\text{imediate reward}} + \underbrace{\gamma \sum_{s'} p(s,a,s') v_\pi(s')}_{\text{expected discounted future value}}$$

BELLMAN EQUATIONS

EXPRESSING VALUE FUNCTIONS IN TERMS OF VALUE FUNCTIONS IN OTHER STATES

$$\underbrace{q_\pi(s,a)}_{\text{value}} = \underbrace{r(s,a)}_{\text{imediate reward}} + \underbrace{\gamma \sum_{s'} p(s,a,s')v_\pi(s')}_{\text{discounted future expected value}}$$

- Recursive expression to compute value functions
- Do not need to consider all possible futures states and rewards *explicitly*

BELLMAN EQUATIONS

MANY FORMS

$$q_\pi(s,a) = \ r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s')$$

BELLMAN EQUATIONS

MANY FORMS

$$q_\pi(s,a) = \; r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s')$$

$$q_\pi(s,a) = \; \sum_{s',r} p(s',r \,|\, s,a)\big(r + \gamma v_\pi(s')\big)$$

BELLMAN EQUATIONS

MANY FORMS

$$q_\pi(s,a) = r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s')$$

$$q_\pi(s,a) = \sum_{s',r} p(s',r \mid s,a)\big(r + \gamma v_\pi(s')\big)$$

$$q_\pi(s,a) = r(s,a) + \gamma \sum_{s'} p(s,a,s') \sum_{a'} \pi(a' \mid s') q_\pi(s',a')$$

BELLMAN EQUATIONS

MANY FORMS

$$v_\pi(s) = \sum_a \pi(a \mid s)\left(r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s') \right)$$

BELLMAN EQUATIONS

MANY FORMS

$$v_\pi(s) = \sum_a \pi(a \mid s)\left(r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s')\right)$$

$$v_\pi(s) = \sum_a \pi(a \mid s) \sum_{s',r} p(s',r \mid s,a)\big(r + \gamma v_\pi(s')\big)$$

OPTIMAL VALUE FUNCTIONS

DEFINITION

$$v_*(s) \doteq \max_{\pi \in \Pi} v_\pi(s)$$

$$q_*(s, a) \doteq \max_{\pi \in \Pi} q_\pi(s, a)$$

OPTIMAL VALUE FUNCTIONS

DEFINITION

v_*, q_* are unique even if there are multiple π_*

- otherwise one would have a larger value than the other and thus one would not be optimal

RECOVER OPTIMAL POLICY

FROM THE OPTIMAL VALUE FUNCTION

Given v_*

$$\pi_*(s) \in \arg\max_a r(s,a) + \gamma \sum_{s'} p(s,a,s') v_*(s')$$

Given q_*

$$\pi_*(s) \in \arg\max_a q_*(s,a)$$

COMPUTE OPTIMAL VALUE

KNOWING OPTIMAL POLICY

$$v_*(s) = \sum_a \pi_*(a \mid s)\left(r(s, a) + \gamma \sum_{s'} p(s, a, s')v_*(s') \right)$$

Still need to solve for v_*

BELLMAN OPTIMALITY EQUATION

DERIVATION

$$
\begin{aligned}
v_*(s) &\doteq \max_{\pi \in \Pi} v_\pi(s) \\
&= \max_{\pi \in \Pi} \sum_a \pi(a \mid s) q_\pi(s, a) \\
&= \max_{\pi \in \Pi} \max_a q_\pi(s, a) \\
&= \max_a q_*(s, a) \\
&= \max_a r(s, a) + \gamma \sum_{s'} p(s, a, s') v_*(s')
\end{aligned}
$$

BELLMAN OPTIMALITY EQUATION

DERIVATION

$$v_*(s) = \max_a r(s, a) + \gamma \sum_{s'} p(s, a, s')v_*(s')$$

FINDING OPTIMAL POLICIES

PROCEDURES

Techniques discussed in next week's module

Computing v_π is a central component

COMPUTING VALUE FUNCTION

USING BELLMAN EQUATION

$$v_\pi(s) = \sum_a \pi(a \mid s) \left(r(s,a) + \gamma \sum_{s'} p(s,a,s') v_\pi(s') \right)$$

$$\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$$

$$v_i = v_\pi(s_i)$$

$$v = [v_1, v_2, \ldots v_n]^\top$$

Write the bellman equation in terms of v_i then solve for all v_i

COMPUTING VALUE FUNCTION

EXPRESSION USING UNKNOWN VALUES

$$v_i = \sum_a \pi(a \mid s) \left(r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j \right)$$

COMPUTING VALUE FUNCTION

EXPRESSION USING UNKNOWN VALUES

$$v_i = \sum_a \pi(a \mid s) \left(r(s, a) + \gamma \sum_j p(s_i, a, s_j) v_j \right)$$

$$v_i = \underbrace{\pi(a_1 \mid s_i) r(s_i, a_1) + \pi(a_2 \mid s_i) r(s_i, a_2) + \ldots}_{= r_i} + \gamma \pi(a_1 \mid s_i) p(s_i, a_1, s_1) v_1 + \gamma \pi(a_1 \mid s_i) p(s_i, a_1, s_2) v_2 + \ldots$$

COMPUTING VALUE FUNCTION

EXPRESSION USING UNKNOWN VALUES

$$v_i = \sum_a \pi(a \mid s)\left(r(s, a) + \gamma \sum_j p(s_i, a, s_j)v_j\right)$$

$$v_i = \underbrace{\pi(a_1 \mid s_i)r(s_i, a_1) + \pi(a_2 \mid s_i)r(s_i, a_2) + \ldots}_{=r_i} + \gamma\pi(a_1 \mid s_i)p(s_i, a_1, s_1)v_1 + \gamma\pi(a_1 \mid s_i)p(s_i, a_1, s_2)v_2 + \ldots$$

$$v_i = r_i + \gamma \underbrace{\sum_a \pi(a \mid s_i)p(s_i, a, s_1)}_{=p_{i,1}} v_1 + \gamma \sum_a \pi(a \mid s_i)p(s_i, a, s_2)v_2 + \ldots \gamma \sum_a \pi(a \mid s_i)p(s_i, a, s_2)v_n$$

COMPUTING VALUE FUNCTION

EXPRESSION USING UNKNOWN VALUES

$$v_i = \sum_a \pi(a \mid s)\left(r(s,a) + \gamma \sum_j p(s_i, a, s_j) v_j \right)$$

$$v_i = \underbrace{\pi(a_1 \mid s_i) r(s_i, a_1) + \pi(a_2 \mid s_i) r(s_i, a_2) + \ldots}_{=r_i} + \gamma \pi(a_1 \mid s_i) p(s_i, a_1, s_1) v_1 + \gamma \pi(a_1 \mid s_i) p(s_i, a_1, s_2) v_2 + \ldots$$

$$v_i = r_i + \gamma \underbrace{\sum_a \pi(a \mid s_i) p(s_i, a, s_1)}_{=p_{i,1}} v_1 + \gamma \sum_a \pi(a \mid s_i) p(s_i, a, s_2) v_2 + \ldots \gamma \sum_a \pi(a \mid s_i) p(s_i, a, s_2) v_n$$

$$v_i = r_i + \gamma p_{i,1} v_1 + \gamma p_{i,2} v_2 + \ldots + \gamma p_{i,n} v_n$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v_1 =&\ r_1 + \gamma p_{1,1} v_1 + \gamma p_{1,2} v_2 + \ldots + \gamma p_{1,n} v_n \\
v_2 =&\ r_2 + \gamma p_{2,1} v_1 + \gamma p_{2,2} v_2 + \ldots + \gamma p_{2,n} v_n \\
\vdots& \\
v_n =&\ r_n + \gamma p_{n,1} v_1 + \gamma p_{n,2} v_2 + \ldots + \gamma p_{n,n} v_n
\end{aligned}
$$

Solve n equations for n unknown variables

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v_1 = &\ r_1 + \gamma p_{1,1} v_1 + \gamma p_{1,2} v_2 + \ldots + \gamma p_{1,n} v_n \\
v_2 = &\ r_2 + \gamma p_{2,1} v_1 + \gamma p_{2,2} v_2 + \ldots + \gamma p_{2,n} v_n \\
\vdots & \\
v_n = &\ r_n + \gamma p_{n,1} v_1 + \gamma p_{n,2} v_2 + \ldots + \gamma p_{n,n} v_n
\end{aligned}
$$

$$
r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \quad P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ldots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{bmatrix}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$v = r + \gamma P v$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v &= r + \gamma P v \\
v - \gamma P v &= r
\end{aligned}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v =& \; r + \gamma P v \\
v - \gamma P v =& \; r \\
(I - \gamma P)v =& \; r
\end{aligned}
$$

$$
I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v &= r + \gamma P v \\
v - \gamma P v &= r \\
(I - \gamma P) v &= r \\
\underbrace{(I - \gamma P)^{-1}(I - \gamma P)}_{A^{-1}A = I} v &= (I - \gamma P)^{-1} r
\end{aligned}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v =&\ r + \gamma P v \\
v - \gamma P v =&\ r \\
(I - \gamma P) v =&\ r \\
\underbrace{(I - \gamma P)^{-1}(I - \gamma P)}_{A^{-1}A = I} v =&\ (I - \gamma P)^{-1} r \\
I v =&\ (I - \gamma P)^{-1} r
\end{aligned}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v =&\ r + \gamma P v \\
v - \gamma P v =&\ r \\
(I - \gamma P) v =&\ r \\
\underbrace{(I - \gamma P)^{-1}(I - \gamma P)}_{A^{-1}A=I}\, v =&\ (I - \gamma P)^{-1} r \\
I v =&\ (I - \gamma P)^{-1} r \\
v =&\ (I - \gamma P)^{-1} r
\end{aligned}
$$

COMPUTING VALUE FUNCTION

SYSTEM OF EQUATIONS

$$
\begin{aligned}
v &= r + \gamma P v \\
v - \gamma P v &= r \\
(I - \gamma P) v &= r \\
\underbrace{(I - \gamma P)^{-1} (I - \gamma P)}_{A^{-1}A = I} v &= (I - \gamma P)^{-1} r \\
I v &= (I - \gamma P)^{-1} r \\
v &= (I - \gamma P)^{-1} r
\end{aligned}
$$

$(I - \gamma P)$ must be invertible (which is not always true, usually when $\gamma = 1$)

In practice, use the Moore-Penrose pseudoinverse if $(I - \gamma P)$ is not invertible

NEXT CLASS

WHAT YOU SHOULD DO

1. Quiz due Wednesday night: Value Functions and Bellman Equations 2

Wednesday: In-class exercises and Q&A