Generalization in Large scale MDPs

Wen Sun

CS 6789: Foundations of Reinforcement Learning

Recap on Bellman Error and Bellman Operator

Two types of Bellman error of $f(s, a)(\approx Q^\star)$

Recap on Bellman Error and Bellman Operator

Two types of Bellman error of $f(s,a)(\approx Q^\star)$

$$BE_Q(s,a) = f(s,a) - \left(r(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)} \max_{a'} f(s',a') \right)$$

Recap on Bellman Error and Bellman Operator

Two types of Bellman error of $f(s,a)(\approx Q^\star)$

$$BE_Q(s,a) = f(s,a) - \left(r(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)} \max_{a'} f(s',a') \right)$$

$$V_f(s) = \arg\max_a f(s,a), \pi_f(s) = \arg\max_a f(s,a)$$

Recap on Bellman Error and Bellman Operator

Two types of Bellman error of $f(s,a)(\approx Q^\star)$

$$BE_Q(s,a) = f(s,a) - \left(r(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)} \max_{a'} f(s',a') \right)$$

$$V_f(s) = \arg\max_a f(s,a), \pi_f(s) = \arg\max_a f(s,a)$$

$$BE_V(s) = V_f(s) - r(s, \pi_f(s)) - \mathbb{E}_{s' \sim P_h(s,\pi_f(s))} V_f(s')$$

Recap on Bellman Error and Bellman Operator

Two types of Bellman error of $f(s,a)(\approx Q^\star)$

$$BE_Q(s,a) = f(s,a) - \left(r(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)} \max_{a'} f(s', a') \right)$$

$$V_f(s) = \arg\max_a f(s,a), \pi_f(s) = \arg\max_a f(s,a)$$

$$BE_V(s) = V_f(s) - r(s, \pi_f(s)) - \mathbb{E}_{s' \sim P_h(s, \pi_f(s))} V_f(s')$$

If $BE(s,a) \neq 0$, then $f \neq Q^\star$

Notations

Probability of π visiting (s, a) at time step h: $d_h^\pi(s, a)$

Question for Today

We have seen tabular MDP and linear MDP, is there a **more general framework** that captures these two, and potentially many more, where efficient learning is possible?

Question for Today

We have seen tabular MDP and linear MDP, is there a **more general framework** that captures these two, and potentially many more, where efficient learning is possible?

In other words, what structural conditions permit RL generalization, provably?

Outline for Today

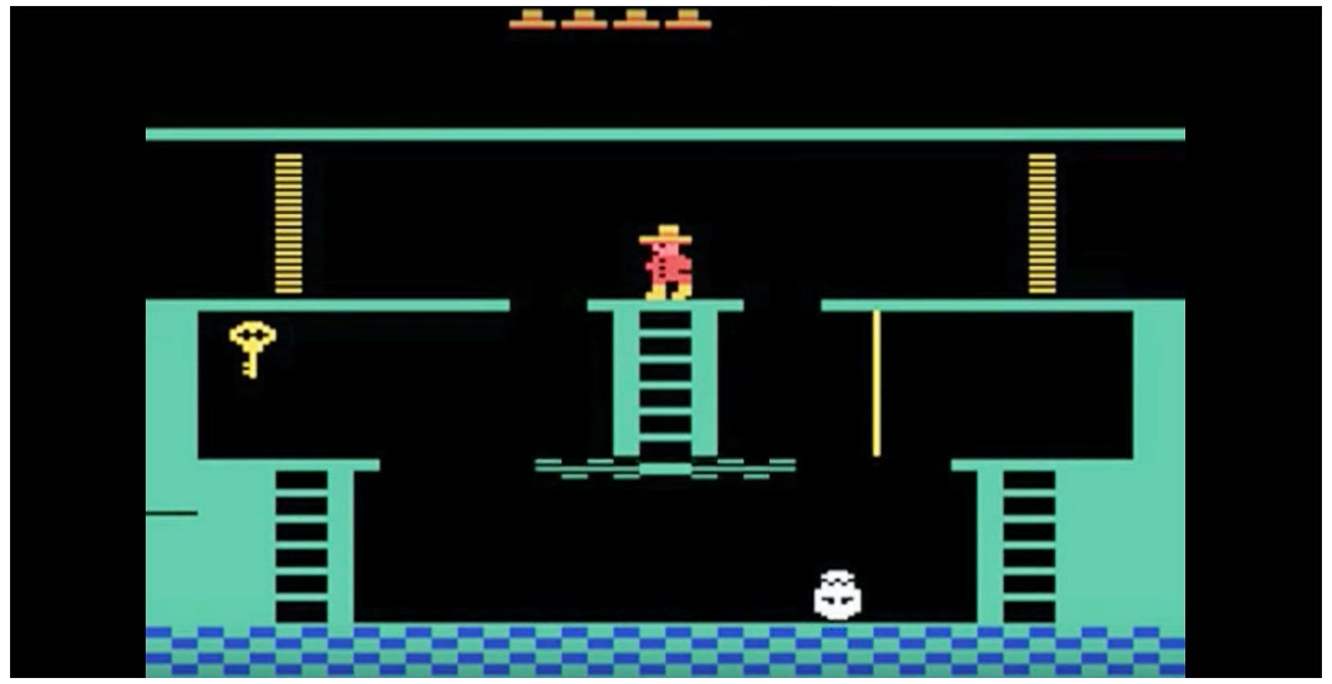
1. Bellman rank Definitions

2. Examples that are captured by the Bellman rank framework

Setting

Finite horizon episodic MDP $\left\{\{S_h\}_{h=0}^{H}, \{A_h\}_{h=0}^{H-1}, H, s_0, r, P\right\}$

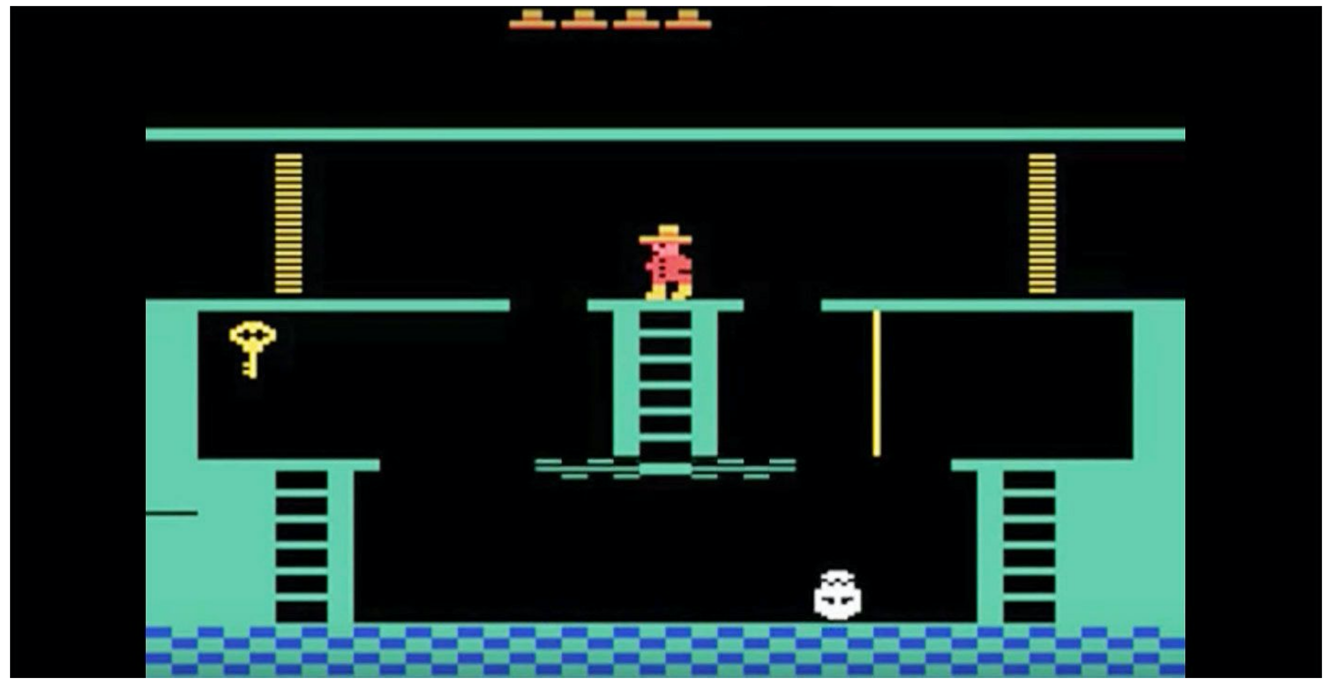
State space S_h is extremely large:

$s =$ 

Setting

Finite horizon episodic MDP $\left\{\{S_h\}_{h=0}^{H}, \{A_h\}_{h=0}^{H-1}, H, s_0, r, P\right\}$

State space S_h is extremely large:

$s =$ 

Not acceptable: poly $\left(|S|\right)$

Need to generalize via (nonlinear) function approximation

Let’s set up function class in RL setting

We will consider **Q function class**

$$\mathcal{F} \subset S \times A \mapsto [0,H]$$

Let's set up function class in RL setting

We will consider **Q function class**

$$\mathcal{F} \subset S \times A \mapsto [0,H]$$

Realizability assumption:

$$Q^\star \in \mathcal{F}$$

Let's set up function class in RL setting

We will consider **Q function class**

$$\mathcal{F} \subset S \times A \mapsto [0,H]$$

Realizability assumption:

$$Q^\star \in \mathcal{F}$$

Define **policy class**: $\Pi = \{\pi : \pi(s) = \arg\max_{a \in A} f(s,a), \forall s \in S \,|\, f \in \mathcal{F}\}$

Let’s set up function class in RL setting

We will consider **Q function class**

$$\mathcal{F} \subset S \times A \mapsto [0,H]$$

Realizability assumption:

$$Q^\star \in \mathcal{F}$$

Define **policy class**: $\Pi = \{\pi : \pi(s) = \arg\max_{a \in A} f(s,a), \forall s \in S \,|\, f \in \mathcal{F}\}$

Define **value function class**: $\mathcal{V} = \{V_f : V_f(s) = \max_{a} f(s,a) \,|\, f \in \mathcal{F}\}$

Learning Goal:

We will do PAC in this lecture rather than regret.

Learning Goal:

We will do PAC in this lecture rather than regret.

Given approximation error ϵ and failure prob δ,
can we learn ϵ *near optimal policy* (i.e., $V^{\hat\pi} \geq V^\star - \epsilon$) in # of samples scaling
poly* with all relevant parameters *(here, we need poly in $\ln(|\mathscr{F}|)$)

How to check if a Q-approximator is good?

We define **average** Bellman error of a Q-estimate g below:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathcal{A}} g(s_{h+1}, a) \right] \right]$$

How to check if a Q-approximator is good?

We define **average** Bellman error of a Q-estimate g below:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

f: defines roll-in distribution over s_h, a_h.

How to check if a Q-approximator is good?

We define **average** Bellman error of a Q-estimate g below:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

f: defines roll-in distribution over s_h, a_h.

We know that $\mathscr{E}(Q^\star; f, h) = 0, \forall f$

How to check if a Q-approximator is good?

We define **average** Bellman error of a Q-estimate g below:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

f: defines roll-in distribution over s_h, a_h.

We know that $\mathscr{E}(Q^\star; f, h) = 0, \forall f$

Hence, any g such that $\mathscr{E}(g; f, h) \neq 0$, is an incorrect $Q^\star$ approximator

How to check if a Q-approximator is good?

We can define **average** Bellman error wrt the V-function induced by g as well:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}}\left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))}\left[V_g(s_{h+1})\right]\right]$$

How to check if a Q-approximator is good?

We can define **average** Bellman error wrt the V-function induced by g as well:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

Again we have $\mathscr{E}(Q^\star; f, h) = 0, \forall f$

How to check if a Q-approximator is good?

We can define **average** Bellman error wrt the V-function induced by g as well:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

Again we have $\mathcal{E}(Q^\star; f, h) = 0, \forall f$

(because: $V_{Q^\star}(s) - r(s, \pi_{Q^\star}(s)) - \mathbb{E}_{s' \sim P_h(.|s, \pi_{Q^\star}(s))} V_{Q^\star}(s') = 0$)

How to check if a Q-approximator is good?

We can define **average** Bellman error wrt the V-function induced by g as well:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

Again we have $\mathscr{E}(Q^\star; f, h) = 0, \forall f$

(because: $V_{Q^\star}(s) - r(s, \pi_{Q^\star}(s)) - \mathbb{E}_{s' \sim P_h(.|s, \pi_{Q^\star}(s))} V_{Q^\star}(s') = 0$)

Hence, any g such that $\mathscr{E}(g; \pi, h) \neq 0$, is an incorrect $Q^\star$ approximator

The Q / V-Bellman rank

$$\forall h : \mathscr{E}_h \in \mathbb{R}^{|\mathscr{F}|\times|\mathscr{F}|}$$

$$\begin{array}{c|c|c|c|c|c|c|}
 & & g & f & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
\pi_f & & \mathscr{E}_{g;f,h} & \mathscr{E}_{f;f,h} & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
\end{array}$$

The Q / V-Bellman rank

$$\forall h : \mathscr{E}_h \in \mathbb{R}^{|\mathscr{F}|\times|\mathscr{F}|}$$

$$\begin{array}{c|c|c|c|c|c|c|}
 & & g & f & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
\pi_f & & \mathscr{E}_{g;f,h} & \mathscr{E}_{f;f,h} & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \\
\hline
\end{array}$$

Rank of this Matrix is defined as Bellman Rank

The Q / V-Bellman rank

In other words, there are two mappings $W_h : \mathcal{F} \mapsto \mathbb{R}^d, \quad X_h : \mathcal{F} \mapsto \mathbb{R}^d$ (d = Bellman-rank)

$$\forall f, g \in \mathcal{F} : \mathcal{E}(g; f, h) = \left\langle W_h(g), X_h(f) \right\rangle$$

Note, we just assume the existence of W, X, but they are unknown

Outline for Today

1. Bellman rank Definitions

2. Examples that are captured by the Bellman rank framework

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s, a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s, a) = r(s, a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s, a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s, a) = r(s, a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

Claim: it has Q-Bellman rank d

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s,a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s,a) = r(s,a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

Claim: it has Q-Bellman rank d

$\forall g(s,a) := \theta^\top \phi(s,a)$, we have:

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s,a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s,a) = r(s,a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

Claim: it has Q-Bellman rank d

$\forall g(s,a) := \theta^\top \phi(s,a)$, we have:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\max_{a \in \mathcal{A}} \theta^\top \phi(s_{h+1}, a) \right] \right]$$

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s,a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s,a) = r(s,a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

Claim: it has Q-Bellman rank d

$\forall g(s,a) := \theta^\top \phi(s,a)$, we have:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} \theta^\top \phi(s_{h+1}, a) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - \mathscr{T}_h(\theta)^\top \phi(s_h, a_h) \right]$$

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s,a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s,a) = r(s,a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s', a'), \forall s, a$$

Claim: it has Q-Bellman rank d

$\forall g(s,a) := \theta^\top \phi(s,a)$, we have:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} \theta^\top \phi(s_{h+1}, a) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - \mathscr{T}_h(\theta)^\top \phi(s_h, a_h) \right]$$

$$= \left\langle \theta - \mathscr{T}_h(\theta), \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} [\phi(s_h, a_h)] \right\rangle$$

The Linear Bellman Completion Model

Given feature ϕ, take any linear function $\theta^\top \phi(s,a)$:

$$\forall h, \exists w \in \mathbb{R}^d, s.t., w^\top \phi(s,a) = r(s,a) + \mathbb{E}_{s' \sim P_h(s,a)} \max_{a'} \theta^\top \phi(s',a'), \forall s,a$$

Claim: it has Q-Bellman rank d

$\forall g(s,a) := \theta^\top \phi(s,a)$, we have:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} \theta^\top \phi(s_{h+1}, a) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[\theta^\top \phi(s_h, a_h) - \mathscr{T}_h(\theta)^\top \phi(s_h, a_h) \right]$$

$$= \left\langle \theta - \mathscr{T}_h(\theta), \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} [\phi(s_h, a_h)] \right\rangle$$

Note linear Bell-completion captures tabular / linear mdp already

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

$$\mathscr{F}_h = \left\{ (w,\theta) : \max_a w^\top \phi(s,a) = \theta^\top \psi(s), \forall s \right\}$$

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

$$\mathscr{F}_h = \left\{ (w, \theta) : \max_a w^\top \phi(s,a) = \theta^\top \psi(s), \forall s \right\}$$

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

$$\mathcal{F}_h = \left\{ (w,\theta) : \max_a w^\top \phi(s,a) = \theta^\top \psi(s), \forall s \right\}$$

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - (w^\star)^\top \phi(s_h, a_h) + \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[(\theta^\star)^\top \psi(s_{h+1}) \right] - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

$$\mathcal{F}_h = \left\{ (w,\theta) : \max_a w^\top \phi(s,a) = \theta^\top \psi(s), \forall s \right\}$$

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - (w^\star)^\top \phi(s_h, a_h) + \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, a_h)} \left[(\theta^\star)^\top \psi(s_{h+1}) \right] - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

$$= \left\langle \begin{bmatrix} w - w^\star \\ \theta - \theta^\star \end{bmatrix}, \quad \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \begin{bmatrix} \phi(s_h, a_h) \\ -\mathbb{E}_{s' \sim P_h(s_h, a_h)}[\psi(s')] \end{bmatrix} \right\rangle$$

The Linear $Q^\star$ & $V^\star$ model:

Assume $Q^\star(s,a) = (w^\star)^\top \phi(s,a), \quad V^\star(s) = (\theta^\star)^\top \psi(s), \forall s, a$

Claim: it has Q-Bellman rank 2d

$$\mathscr{F}_h = \left\{ (w,\theta) : \max_a w^\top \phi(s,a) = \theta^\top \psi(s), \forall s \right\}$$

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h,a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

$$= \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[w^\top \phi(s_h, a_h) - (w^\star)^\top \phi(s_h, a_h) + \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h,a_h)} \left[(\theta^\star)^\top \psi(s_{h+1}) \right] - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h,a_h)} \left[\theta^\top \psi(s_{h+1}) \right] \right]$$

$$= \left\langle \begin{bmatrix} w - w^\star \\ \theta - \theta^\star \end{bmatrix}, \quad \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \begin{bmatrix} \phi(s_h, a_h) \\ -\mathbb{E}_{s' \sim P_h(s_h,a_h)}[\psi(s')] \end{bmatrix} \right\rangle$$

As we will see, linear Q*&V* is learnable, and recall linear Q* is not…

$Q^\star$ - state abstraction

We have a small latent state space Z, and a **known** mapping ξ from state s to z

$$Q^\star(s_1, a) = Q^\star(s_2, a), \forall a, \text{if } \xi(s_1) = \xi(s_2)$$

$Q^\star$ - state abstraction

We have a small latent state space Z, and a **known** mapping ξ from state s to z

$$Q^\star(s_1, a) = Q^\star(s_2, a), \forall a, \text{if } \xi(s_1) = \xi(s_2)$$

Claim: this model has Q-Bellman rank $|Z||A| + |Z|$

We can show that this model is captured by linear $Q^\star$ & $V^\star$

Low-rank MDP

$$P_h(s' \mid s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Low-rank MDP

$P_h(s' \mid s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Low-rank MDP

$$P_h(s' \mid s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Define representation class Φ, with $\phi^\star \in \Phi$

$$\mathscr{F}_h = \{\theta^\top \phi(\cdot, \cdot) : \|\theta\|_2 \leq W, \phi \in \Phi\}$$

Low-rank MDP

$$P_h(s' | s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Define representation class Φ, with $\phi^\star \in \Phi$

$$\mathscr{F}_h = \{\theta^\top \phi(\cdot, \cdot) : \|\theta\|_2 \leq W, \phi \in \Phi\}$$

$$\mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot | s_h, \pi_g(s_h))}[V_g(s_{h+1})] \right]$$

Low-rank MDP

$P_h(s' \mid s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Define representation class Φ, with $\phi^\star \in \Phi$

$$\mathscr{F}_h = \{\theta^\top \phi(\cdot,\cdot) : \|\theta\|_2 \le W, \phi \in \Phi\}$$

$$\mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})] \right]$$

$$= \mathbb{E}_{\tilde{s}, \tilde{a} \sim d_{h-1}^{\pi_f}} \mathbb{E}_{s_h \sim P_{h-1}(\cdot|\tilde{s}, \tilde{a})} \left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})] \right]$$

Low-rank MDP

$P_h(s'|s,a) = \mu_h^\star(s')^\top \phi_h^\star(s,a)$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Define representation class Φ, with $\phi^\star \in \Phi$

$$\mathscr{F}_h = \{\theta^\top \phi(\cdot,\cdot) : \|\theta\|_2 \leq W, \phi \in \Phi\}$$

$$\mathbb{E}_{s_h \sim d_h^{\pi_f}}\left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1}\sim P_h(\cdot|s_h,\pi_g(s_h))}[V_g(s_{h+1})]\right]$$

$$= \mathbb{E}_{\tilde{s},\tilde{a}\sim d_{h-1}^{\pi_f}} \mathbb{E}_{s_h \sim P_{h-1}(\cdot|\tilde{s},\tilde{a})}\left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1}\sim P_h(\cdot|s_h,\pi_g(s_h))}[V_g(s_{h+1})]\right]$$

$$= \mathbb{E}_{\tilde{s},\tilde{a}\sim d_{h-1}^{\pi_f}} \int_{s_h} \mu_{h-1}^\star(s_h)^\top \phi_{h-1}^\star(\tilde{s},\tilde{a})\left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1}\sim P_h(\cdot|s_h,\pi_g(s_h))}[V_g(s_{h+1})]\right] d(s_h)$$

Low-rank MDP

$$P_h(s' \mid s, a) = \mu_h^\star(s')^\top \phi_h^\star(s, a)$$ (neither $\mu^\star$ nor $\phi^\star$ is known)

Claim: this model has V-Bellman rank d

Define representation class Φ, with $\phi^\star \in \Phi$

$$\mathcal{F}_h = \{\theta^\top \phi(\cdot,\cdot) : \|\theta\|_2 \le W, \phi \in \Phi\}$$

$$\mathbb{E}_{s_h \sim d_h^{\pi_f}}\left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})]\right]$$

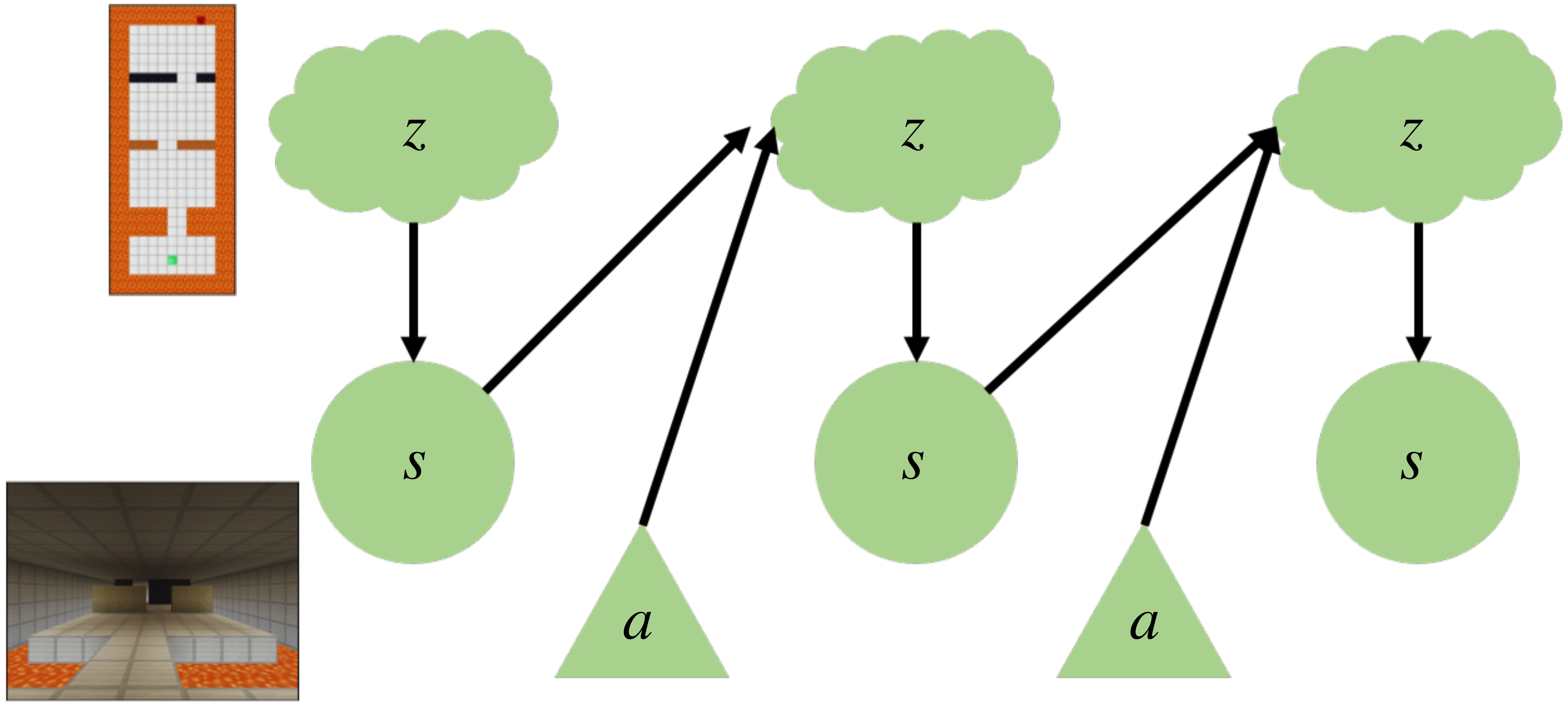
$$= \mathbb{E}_{\tilde{s},\tilde{a} \sim d_{h-1}^{\pi_f}} \mathbb{E}_{s_h \sim P_{h-1}(\cdot|\tilde{s},\tilde{a})}\left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})]\right]$$

$$= \mathbb{E}_{\tilde{s},\tilde{a} \sim d_{h-1}^{\pi_f}} \int_{s_h} \mu_{h-1}^\star(s_h)^\top \phi_{h-1}^\star(\tilde{s},\tilde{a}) \left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})]\right] d(s_h)$$

$$= \left\langle \int_{s_h} \mu_{h-1}^\star(s_h) \left[V_g(s_h) - r(s, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P_h(\cdot|s_h, \pi_g(s_h))}[V_g(s_{h+1})]\right] d(s_h), \quad \mathbb{E}_{\tilde{s},\tilde{a} \sim d_{h-1}^{\pi_f}}[\phi_{h-1}^\star(\tilde{s},\tilde{a})] \right\rangle$$

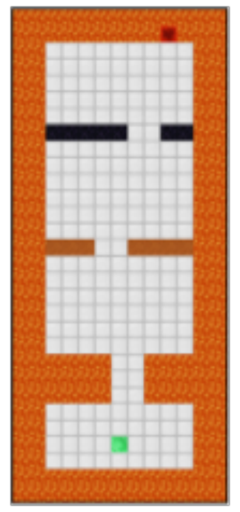
Latent variable MDP

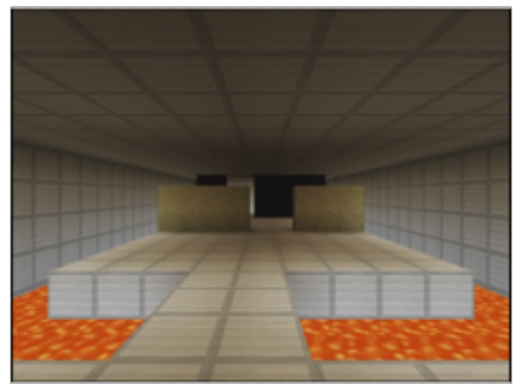
Latent variable MDP is captured by low-rank MDP, so it has small V-Bellman rank…

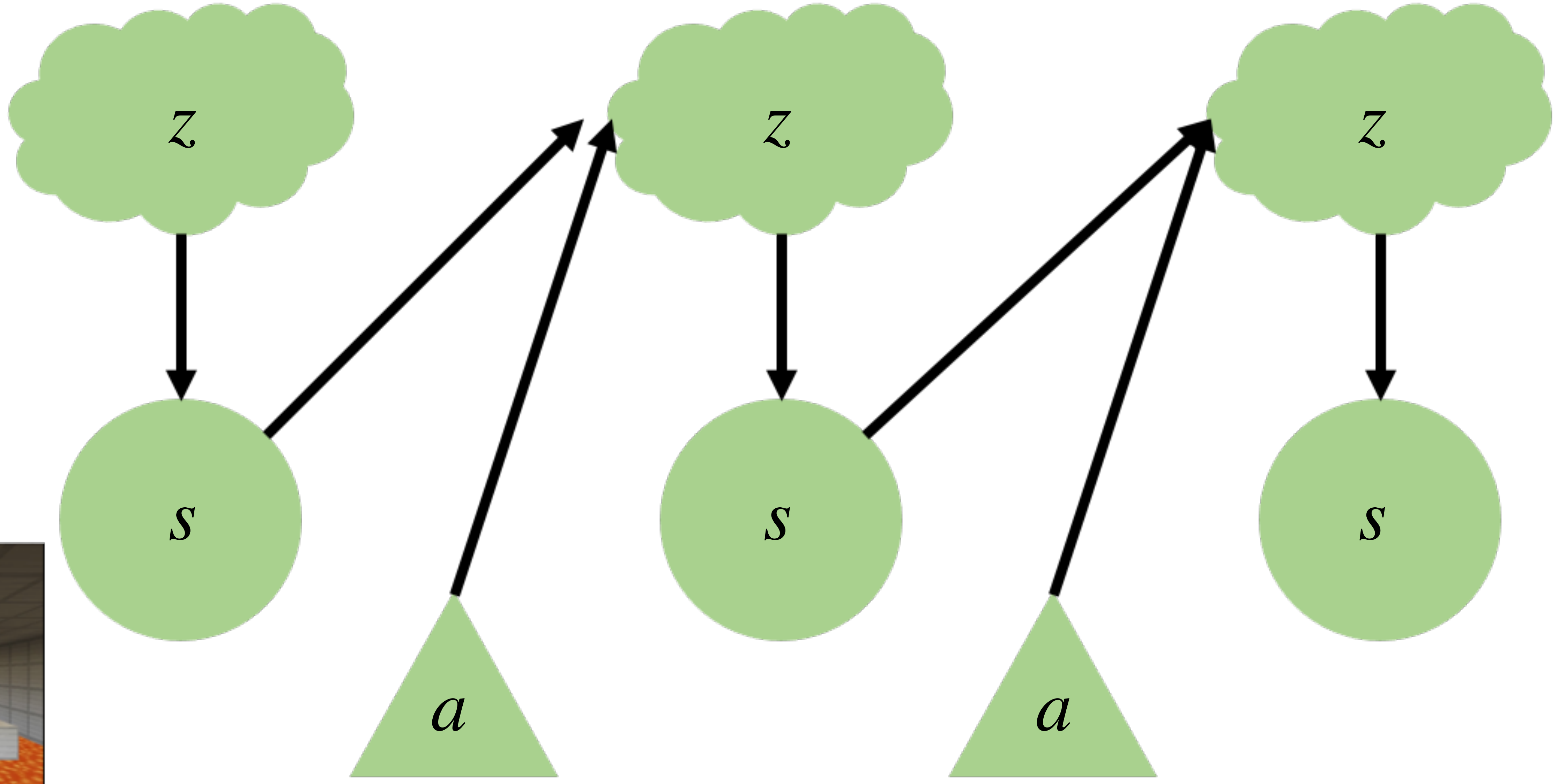


Latent variable MDP

Latent variable MDP is captured by low-rank MDP, so it has small V-Bellman rank…



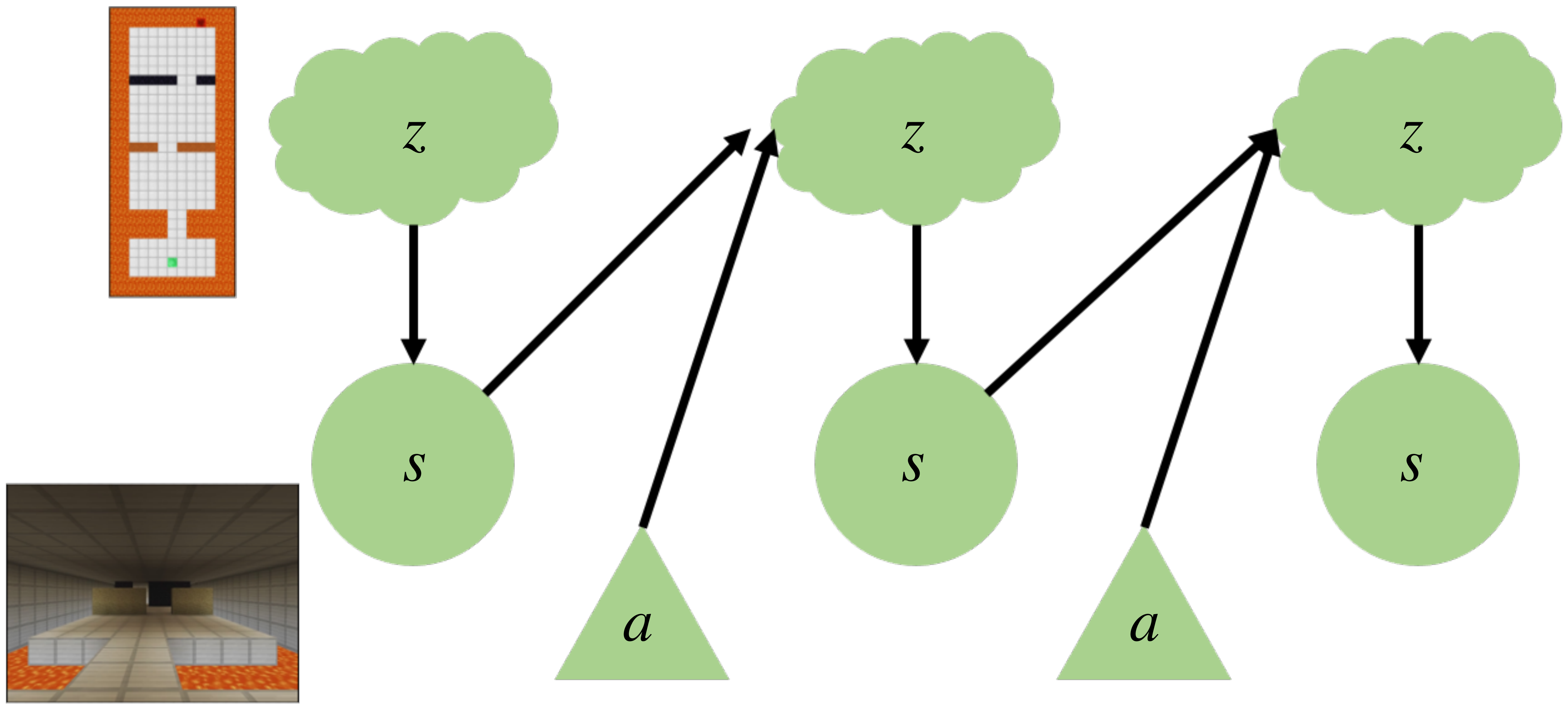




Given s, a: $z \sim \phi^\star(s, a), s' \sim \nu^\star(z)$

Latent variable MDP

Latent variable MDP is captured by low-rank MDP, so it has small V-Bellman rank…



Given s, a: $z \sim \phi^\star(s,a), s' \sim \nu^\star(z)$

V-Bellman rank = Number of latent states

Summary

1. Q-Bellman rank: related to the Bellman error of a Q function estimate g:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

Summary

1. Q-Bellman rank: related to the Bellman error of a Q function estimate g:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot|s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

2. V-Bellman rank: related to the Bellman error of a V function estimate

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot|s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

Summary

1. Q-Bellman rank: related to the Bellman error of a Q function estimate g:

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathscr{A}} g(s_{h+1}, a) \right] \right]$$

2. V-Bellman rank: related to the Bellman error of a V function estimate

$$\mathscr{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

3. Small Bellman rank means that:

$$\forall f, g \in \mathscr{F} : \mathscr{E}(g; f, h) = \left\langle W_h(g), X_h(f) \right\rangle$$

where $X_h(f), W_h(f)$ are low-dim vectors

Summary

1. Q-Bellman rank: related to the Bellman error of a Q function estimate g:

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h, a_h \sim d_h^{\pi_f}} \left[g(s_h, a_h) - r(s_h, a_h) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, a_h)} \left[\max_{a \in \mathcal{A}} g(s_{h+1}, a) \right] \right]$$

2. V-Bellman rank: related to the Bellman error of a V function estimate

$$\mathcal{E}(g; f, h) = \mathbb{E}_{s_h \sim d_h^{\pi_f}} \left[V_g(s_h) - r(s_h, \pi_g(s_h)) - \mathbb{E}_{s_{h+1} \sim P(\cdot | s_h, \pi_g(s_h))} \left[V_g(s_{h+1}) \right] \right]$$

3. Small Bellman rank means that:

$$\forall f, g \in \mathcal{F} : \mathcal{E}(g; f, h) = \left\langle W_h(g), X_h(f) \right\rangle$$

where $X_h(f), W_h(f)$ are low-dim vectors

4. Many models (more in the book chapter) indeed have low-Q or V Bellman rank

Next week:

A general algorithm that can learn an ϵ near optimal policy w/ # of samples

$$\text{poly}(H, 1/\epsilon, \ln(|\mathscr{H}|), \text{b-rank})$$